REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1990	XXXXX/DISSERTATION

4. TITLE AND SUBTITLE
Analysis of Multivariable Control Systems in the Presence of Structured Uncertainties

5. FUNDING NUMBERS

6. AUTHOR(S)
Robert Jeffrey Norris

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
AFIT Student Attending: University of Florida

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/CI/CIA-90-026D

9. SPONSORING, MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFIT/CI
Wright-Patterson AFB OH 45433-6583

10. SPONSORING, MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release IAW 190-1
Distributed Unlimited
ERNEST A. HAYGOOD, 1st Lt, USAF
Executive Officer

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

DTIC
ELECTE
NOV 02 1990
S D
B

14. SUBJECT TERMS

15. NUMBER OF PAGES
133

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

AD-A227 976

DTIC FILE COPY

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1 Agency Use Only *(Leave blank)*

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	PR	- Project
G	- Grant	TA	- Task
PE	- Program Element	WU	- Work Unit Accession No.

Block 6. Author(s) Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9 Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as. Prepared in cooperation with. ., Trans. of..., To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- **DOD** - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- **DOE** - See authorities.
- **NASA** - See Handbook NHB 2200.2.
- **NTIS** - Leave blank.

Block 12b. Distribution Code.

- **DOD** - Leave blank.
- **DOE** - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- **NASA** - Leave blank.
- **NTIS** - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

026D

ANALYSIS OF MULTIVARIABLE CONTROL SYSTEMS IN THE PRESENCE OF STRUCTURED UNCERTAINTIES

By

ROBERT JEFFREY NORRIS

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1990

90 11 [illegible] 073

To

Diana, Sarah, Richard

and

Angela

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor, Dr. Latchman, for his guidance and encouragement throughout the course of my studies. Dr. Latchman has always found time to discuss my work and for that I am especially thankful.

I also wish to thank the other members of my committee for their advice and help on a number of different topics. I am indebted to Dr. Bullock for his many hours of assistance with the computer systems, especially concerning LaTeXand the laser printers. I wish to thank Dr. Sigmon for his help with MATLAB^{TM} and with numerous questions concerning linear algebra. I would also like to thank Dr. Zimmerman for his advice on numerical optimization algorithms which greatly simplified the implementation of my results. Finally, I wish to thank Dr. Basile for his help during the early stages of my research, and Dr. Carroll for graciously agreeing to serve on my committee at very short notice.

A number of my fellow students have provided both inspiration and advice. In particular, I wish to thank José Letra, Peter Young, and Shannon Fields for their help over the past three years.

A final acknowledgement goes to the United States Air Force for this generous opportunity to continue my education.

DTIC COPY INSPECTED

By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ANALYSIS OF MULTIVARIABLE CONTROL SYSTEMS IN THE PRESENCE OF STRUCTURED UNCERTAINTIES

By

ROBERT JEFFREY NORRIS

August 1990

Chairman: Dr. Haniph A. Latchman
Major Department: Electrical Engineering

An analysis of the stability properties of uncertain multivariable control systems in the frequency domain is presented. Necessary and sufficient stability criteria are reviewed along with singular value scaling techniques for characterizing permissible uncertainties. Such scaling methods have become widely accepted tools for the analysis of control systems in the presence of structured uncertainties. Included in this study are the general block similarity scaling techniques advanced by Doyle and the nonsimilarity scaling approach of Kouvaritakis and Latchman.

For element-by-element structured uncertainties, both scaling methods reliably compute Doyle's structured singular value, μ, which provides an indication of system stability. However, the similarity scaling formulation has the disadvantage of expanding an $n \times n$ system matrix to an $n^2 \times n^2$ matrix requiring $n^2 - 1$ optimization variables to compute μ. Using nonsimilarity scaling, the system size remains n with the additional benefit of requiring only $2(n-1)$ optimization variables.

over

The results of this work show that for scalar uncertainties, the structure may be exploited to yield a similarity scaling method which requires no more than the $2(n-1)$ optimization variables needed for nonsimilarity scaling. Substantial savings in floating point operations are observed for various system sizes enhancing the capability of this method for analysis and iterative design. A similar reduction in optimization variables is shown to hold for the important class of general block structured uncertainties. This reduction leads to a complete solution for the 2×2 block uncertainty problem.

In addition, a direct relationship between the similarity and nonsimilarity scaling matrices is presented. This direct relationship, coupled with a reduction of optimization variables like that shown for similarity scaling, provides a more efficient implementation of the Fan and Tits vector optimization method for computing μ.

For systems with repeated maximum singular values, both similarity and nonsimilarity scaling procedures generally fail to calculate μ exactly. By invoking the major principal direction alignment principle of Kouvaritakis and Latchman, a $2(q-1)$-dimensional optimization problem is proposed for estimating μ where q represents the multiplicity of the maximum singular value. In all numerical experience to date, these estimates and the corresponding exact values of μ have agreed within three significant figures. Application of these techniques to design is discussed.

CHAPTER 1
INTRODUCTION TO SYSTEM UNCERTAINTY

1.1 Introduction

The first task in any control design is the modelling process through which a mathematical representation of the system is developed. This model may be constructed theoretically through some knowledge of the physical laws involved or empirically by characterizing experimental data collected over the range of operating conditions [1]. Theoretical models may assume linearity around some nominal operating point while neglecting nonlinear factors that may dominate away from this nominal point. Similarly, the range of operating conditions chosen for the empirical model may not be sufficiently exhaustive to adequately characterize system behavior. For both modelling approaches, the effects of component aging, temperature and pressure variations, manufacturing tolerances and countless other unknown factors combine to cast doubt on the accuracy of a particular system model. It is therefore critical to ensure that a controller designed for a nominal model does not lose required stability and performance properties when applied to the real-world system [2].

For Single-Input Single-Output (SISO) systems, model uncertainty problems have typically been addressed by ensuring that adequate gain and phase margins exist throughout the range of operating conditions. These margins could then absorb the detrimental effects of the uncertainties without sacrificing stability requirements.

For Multi-Input Multi-Output (MIMO) systems, the classical definitions of gain and phase margins do not apply because of complex system interactions [3]. Examples of multivariable systems include chemical plants, nuclear facilities and high performance aircraft where unstable operation may result in the loss of expensive equipment or even lives [4, 5].

This dissertation reviews some of the current frequency domain techniques employed in the analysis of multivariable systems with uncertain plant models. Of particular importance are those approaches that address structured uncertainties because they generally produce a much less conservative stability analysis than those dealing strictly with unstructured uncertainties. Alternative formulations of these approaches are developed offering substantial improvements in computational efficiency.

Following this introduction is a brief discussion of the historical treatment of uncertainties for feedback systems as well as a list of standard notation used throughout this work. Chapter 2 summarizes the classical Nyquist stability criterion for SISO systems as well as the generalized nyquist criterion for MIMO systems. Singular value techniques that provide sufficient stability criteria are reviewed in Chapter 3 while Chapter 4 covers various scaling techniques that recover the necessity of the stability criteria under most conditions. The main contributions of this dissertation appear in Chapters 5, 6, 7, and 8 along with several examples that illustrate the advantages offered by these new results. Finally, Chapter 9 summarizes the work and introduces some of the promising directions for future robust control research.

1.2 Historical Treatment of Uncertainty

One of the first efforts to account for model uncertainties during control system design was by H. S. Black of Bell Laboratories in 1927 [6]. Black introduced the concept of feedback to eliminate amplifier distortion in long distance telephone communications. Amplifiers of that time required hourly adjustments of bias currents resulting in unacceptable manpower demands [7]. Black's feedback amplifier proved to be almost completely immune to amplifier uncertainties caused by nonlinearities and temperature and aging changes. One drawback to the new amplifier design was an occasional self-oscillation or "singing" at certain loop gain settings. In response to this sometimes severe problem, another Bell Laboratories scientist, H. Nyquist, developed the now famous Nyquist stability criterion relating closed loop system stability to open loop frequency response information [8]. From this stability theory came the pioneering work of H. W. Bode concerning the issue of stability robustness in controller design [9]. The resulting concepts of gain and phase margins for SISO systems provided the design engineer with an measure of stability in the presence of uncertainties.

The Nyquist stability criterion is recognized as a major advance in the design of stable SISO control systems and it forms the basis of classical control theory. However, events of the 1950s and 1960s shifted the emphasis of control theory research from the frequency domain to the time domain representation as worldwide attention focused on manned and unmanned rocket guidance and control. With almost unlimited budgets and well defined models, the problems associated with system un-

certainty became (at least temporarily) less important. However, the late 1970s saw renewed interest in frequency domain analysis and design as the precise plant models required for optimal state-space methods simply could not be determined for many important systems due to numerous plant uncertainties [10].

Finally, in 1977 a multivariable analogue to the Nyquist stability criterion known as the characteristic loci method (CLM) was developed [11]. Introduced by MacFarlane and Postlethwaite, the CLM utilizes frequency dependent plots of transfer matrix eigenvalues (called loci) to characterize MIMO system stability. In its original formulation, the CLM does not provide easily discernable information on stability margins because slight perturbations in the transfer matrix can cause large shifts in the loci. This difficulty led to the use of singular value techniques as a means of extending the CLM to account for system uncertainty.

For unstructured uncertainties, which presume no knowledge of an uncertainty other than a norm bound, Doyle and Stein have shown that singular value bounds can provide necessary and sufficient stability conditions for uncertain multivariable systems [12]. Unfortunately, this formulation fails to take advantage of any available knowledge concerning the possibly well-defined structure of the uncertainty. The controller must then accommodate uncertainties that may be physically impossible leading to an overly conservative design [13].

Safonov applied the concept of similarity scaling to singular value analysis for the case of diagonally structured scalar uncertainties [14]. This scaling idea was then extended by Doyle to consider the important class of general block structured

perturbations which allow norm bounded uncertainty blocks to be of arbitrary dimension [15]. Doyle's concept of the "structured singular value" (denoted μ) as a means of determining the set of permissible structured uncertainties for system stability has become a widely accepted tool in the design of robust, multivariable control systems. Additional scaling techniques such as the nonsimilarity approach of Kouvaritakis and Latchman have evolved to address element- by-element structured uncertainties and, in fact, both similarity and nonsimilarity scaling methods have been shown to produce necessary and sufficient stability conditions in terms of μ for most systems with structured uncertainties [16, 17].

These singular value techniques have recently been extended to the areas of H^∞ and μ-Synthesis: both of which allow a great deal of flexibility in the satisfaction of controller design requirements while retaining stability and performance robustness properties [18, 19]. The main emphasis of this work will therefore be directed towards the singular value analysis techniques including new results that enhance their application in the analysis of robust multivariable control systems.

1.3 Notation

The following notational convention will apply unless otherwise stated.

$C^{n\times m}$: The set of complex matrices with n rows and m columns.

$\Re^{n\times m}$: The set of real matrices with n rows and m columns.

$\jmath$: The square root of -1.

$Im(a)$: The imaginary portion of complex element a.

$Re(a)$: The real portion of complex element a.

$\arg(a)$: The argument of complex element a.

$|a|$: The absolute value of element a.

$\overline{a}$: The complex conjugate of scalar a.

A^H : The complex conjugate transpose of matrix A.

$\|A\|_p$: The p-norm of matrix A ($p = 2$ unless noted otherwise).

$\|A\|_F$: The Frobenius norm of matrix A.

A^+ : Matrix A with elements replaced with their absolute values.

A^{-1} : Inverse of matrix A.

$det\{A\}$: The determinant of matrix A.

$\lambda_i(A)$: The ith eigenvalue of matrix A.

$\rho(A)$: The spectral radius of A.

$\sigma_i(A)$: The ith singular value in magnitude.

$\overline{\sigma}(A)$: The singular value with maximum magnitude ($\overline{\sigma} = \sigma_1$).

$\mu(A)$: The structured singular value of matrix A.

$\mathcal{D}$: The family of diagonal matrices with positive, real entries.

$\mathcal{U}$: The family of diagonal unitary matrices.

inf : Infimum.

sup : Supremum.

max : Maximum.

■ : The completion of a proof or discussion.

CPU : Central processing unit.

CHAPTER 2
FREQUENCY DOMAIN ANALYSIS

2.1 Transfer Matrix Representation

Any real world system can be characterized by a relationship between system inputs and outputs. Given a system with output $y(s)$ and input $u(s)$, the transfer function $G(s)$ relates the two in the manner

$$y(s) = G(s)u(s).$$

For nonscalar $G(s)$, the off-diagonal elements of $G(s)$ produce the system interaction that complicates multivariable control systems. The identification process to determine a transfer matrix begins by injecting known inputs $u(s)$ into the plant and measuring the resulting output $y(s)$. An alternative approach begins with the time domain state space equations usually derived through some knowledge of the plant's physics. The state space representation may be expressed as

$$\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t)$$

where A, B, C, and D are real, possibly time varying matrices. Through the Laplace transform, the unique transfer matrix representation may then be written as

$$G(s) = C(sI - A)^{-1}B + D.$$

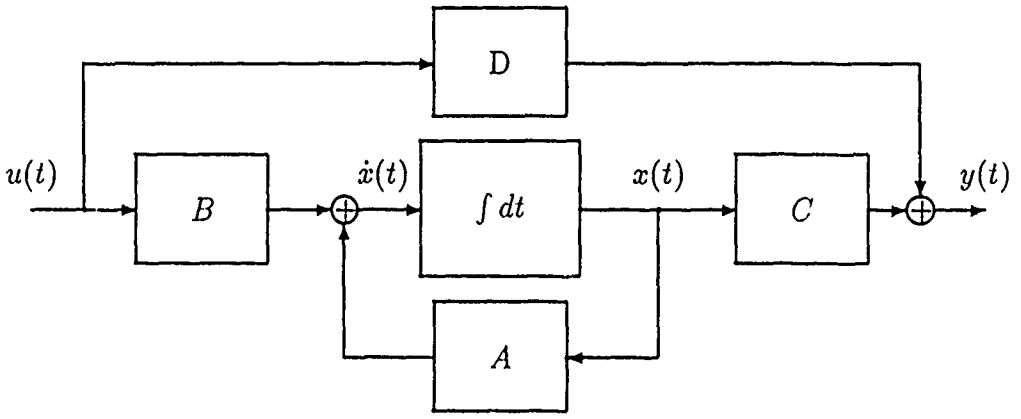

Figure 2.1: State Space System Representation.

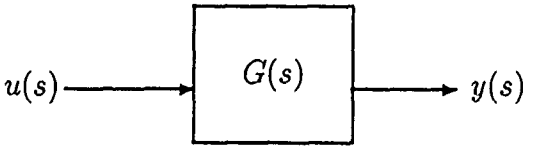

Figure 2.2: Transfer Matrix System Representation.

2.2 Characteristic Loci and the Generalized Nyquist Criterion

Two assumptions are made at this point concerning the transfer matrices used throughout this work. The first assumption is that $G(s)$ is square and rational. The second assumption is that $G(s)$ contains no hidden unstable modes. Also, only unity feedback will be addressed explicitly although any uncompensated $G(s)$ may simply be rewritten as $\hat{G}(s) = K(s)G(s)$ where $K(s)$ is some precompensator. The formulation would then continue with $\hat{G}(s)$ instead of $G(s)$.

Denote $g(s) = \frac{n(s)}{d(s)}$ as the open-loop transfer function of a scalar system where $n(s)$ and $d(s)$ represent the numerator and denominator polynomials respectively. Then $g(s)$ may be related to the return difference transfer function, $f(s)$, by

$$f(s) = 1 + g(s) = \frac{n(s) + d(s)}{d(s)}. \tag{2.1}$$

The term $n(s) + d(s)$ from Equation 2.1 actually corresponds to the closed-loop pole polynomial, $p_c(s)$, under unity feedback while $d(s)$ corresponds to the open-loop pole polynomial, p_o, of $g(s)$. Therefore $f(s)$ may be rewritten as the ratio of the closed-loop to open-loop pole polynomials or

$$f(s) = 1 + g(s) = \frac{p_c(s)}{p_o(s)}. \tag{2.2}$$

Application of the principle of the argument allows the development of a graphical stability test for Equation 2.2. Denote the open and closed-loop right-half plane poles as n_o and n_c respectively. Mapping $g(s)$ onto the classical Nyquist D-contour requires that the number of clockwise critical point $(-1 + j0)$ encirclements equals $n_c - n_o$. Closed-loop stability implies that $n_c = 0$ so the number of critical point encirclements

for a stable system must be $-n_o$. This result forms the basis for the classical Nyquist stability criterion.

Extending these results to the multivariable case, MacFarlane and Postlethwaite generalized Equation 2.2 to the form

$$det\{F(s)\} = det\{[I + G(s)]\} = \alpha \frac{p_c(s)}{p_o(s)}$$

where $F(s)$ is the multivariable return difference matrix and α is a scalar constant [11]. Rather than mapping $g(s)$ directly as in the SISO case, the characteristic loci, defined as the Nyquist D-contour map of the frequency dependent eigenfunctions, $g_i(s)$, are mapped. These eigenfunctions are solutions to the equation

$$det\{[g_i(s)I - G(s)]\} = 0$$

and hence form algebraic functions of the elements of $G(s)$. Also, the $g_i(s)$ are analytic everywhere except where two or more of the $g_i(s)$ coincide. At such points, the separate layers of the Riemann surface which forms the domain of the $g_i(s)$ are simply joined together forming closed curves. Once again the principle of the argument may be applied to these closed curves to form the basis of the generalized Nyquist criterion stated in Theorem 1.1 [11].

Theorem 1.1: A multivariable system with open-loop transfer matrix $G(s)$ containing only controllable and observable modes will be stable under unity feedback if and only if the net sum of counter-clockwise characteristic loci encirclements of the critical point $(-1 + j0)$ is equal to the number of right-half plane open-loop poles of $G(s)$.

This important theorem allows for a complete graphical stability test of multivariable systems in the frequency domain where no model uncertainty is present. Unfortunately, SISO concepts of system robustness based on gain and phase margins do not extend to the multivariable case as the eigenfunctions may display large shifts from relatively small perturbations in the elements of $G(s)$. The next section addresses the stability of systems in the presence of uncertainties.

2.3 Necessary and Sufficient Stability Conditions

The generalized Nyquist criterion, developed from the characteristic loci method, provides both necessary and sufficient stability conditions for the nominal plant denoted $G_o(s)$ where $s = j\omega$ on the Nyquist D contour. Stability is guaranteed if and only if the number of counter-clockwise critical point encirclements by the characteristic loci equals the number of unstable open loop poles. Assuming that the nominal plant itself is stable (either alone or following some type of control implementation), instability in the perturbed plant, $G_p(s)$, implies a change in the number of critical point encirclements. If $G_o(s)$ and $G_p(s)$ have the same number of of open-loop unstable poles, then for such a change to occur, at least one eigenvalue of $G_p(s)$ must pass through the critical point $(-1 + j0)$. Restricting the nominal and perturbed plants to the same number of open-loop poles requires uncertainty matrix Δ to be stable transfer matrix. While this requirement does limit allowable uncertainty structures to those with no poles in the right-half plane, Foo has shown that unstable perturbations can generally be decomposed into two stable perturbations allowing the analysis to continue with a possible increase in conservatism [20].

For additive uncertainties, the necessary and sufficient stability criterion for an uncertain plant $G_p(s) = (G_o(s)+\Delta(s))$, where $\Delta(s)$ represents a frequency dependent uncertainty matrix, becomes

$$\lambda(G_p(s)) \neq -1.$$

Substituting back the nominal plant and $\Delta(s)$, the criterion may be written as

$$\lambda(G_o(s) + \Delta(s)) \neq -1$$

or

$$\lambda(G_o(s) + \Delta(s) + I) \neq 0. \tag{2.3}$$

For any matrix A, the determinant of A equals the product of the eigenvalues of A so Equation 2.3 becomes

$$det\{(G_o(s) + \Delta(s) + I)\} \neq 0$$

or

$$det\{(G_o(s) + I\} \cdot det\{(I + G_o(s))^{-1}\Delta(s) + I\} \neq 0. \tag{2.4}$$

The nominal plant is known to be stable and hence cannot have any eigenvalues passing through the critical point. Applying this fact, Equation 2.4 may be simplified to the form

$$\lambda(I + M(s)\Delta(s)) \neq 0 \qquad \text{where } M(s) = (I + G_o(s))^{-1}$$

or

$$\lambda(M(s)\Delta(s)) \neq -1. \tag{2.5}$$

Typically, some form of norm bound on either the entire matrix (unstructured uncertainties) or individual elements (structured uncertainties) can be estimated, leaving the phase information to vary freely. This freedom in choosing the phase of Δ allows Equation 2.5 to be written as the necessary and sufficient stability condition

$$\sup_{\Delta(s)} \rho(M(s)\Delta(s)) < 1 \tag{2.6}$$

where ρ denotes the spectral radius. It should be noted that while the preceding argument addresses additive uncertainties, a similar approach can be applied to multiplicative uncertainties where $G_p(s) = G_o(s)(I + \Delta(s))$. Multiplying on the right by $G_o(s)$ gives

$$G_p(s) = G_o(s) + G_o(s)\Delta(s)$$

or

$$G_p(s) = G_o(s) + \hat{\Delta}(s) \quad \text{where } \hat{\Delta}(s) = G_o(s)\Delta(s).$$

Although Equation 2.6 provides necessary and sufficient stability conditions for uncertain systems, it is quite difficult to compute because the entire range of Δ must be considered. Furthermore, the solution lacks convexity allowing for the possibility of multiple maxima. This unfortunate situation greatly complicates attempts to find the true supremum of $\rho(M(s)\Delta(s))$ over $\Delta(s)$. Because of the difficulties involved in solving Equation 2.6, singular value techniques were developed to provide upper bounds on permissible model uncertainties. Before introducing these techniques, a mathematical description of several important uncertainty classes will be presented.

2.4 Uncertainty Classes

So far, the frequency dependent uncertainty matrix, $\Delta(s)$, has been used to represent the additive or multiplicative uncertainty present in control system. No information has been provided on the structure of $\Delta(s)$ which turns out to be quite important in the actual mathematical analysis of system stability. Rather than proposing one all-encompassing model for $\Delta(s)$ to account for system uncertainty, three representations will be presented with varying degrees of information required for each one. In developing these three models, two main objectives are observed:

1. The model should handle all information available concerning the system's actual uncertainties. All impossible uncertainty structures should be excluded to reduce conservatism of the stability analysis.

2. The model should be as simple as possible so that the analysis process is not unnecessarily complex. Simplified models also encourage an interactive design process so various control designs may be quickly generated and compared.

Naturally, these objectives produce conflicting requirements and careful consideration must be used to select an appropriate model. In any event, a conservative stability analysis is preferable over one that is simple but erroneous.

The three uncertainty models considered here include unstructured uncertainties, element-by-element structured uncertainties, and block diagonal structured uncertainties. In each of these classes, the uncertainty matrix $\Delta(s)$ is complex with some form of norm bound placed on the magnitude of the matrix or matrix elements. The

simplest uncertainty class to describe and manipulate concerns unstructured uncertainties which may be mathematically represented as

$$D_u = \{\Delta \mid \|\Delta\|_2 \leq \delta \in \Re^+\}$$

where $\|\cdot\|_2$ represents the standard matrix 2-norm and δ a real scalar. (Here, and in the remainder of this work, frequency dependence will be implied). From the definition, it is apparent that no information concerning the inner structure of Δ is accessible. This effectively places a SISO bound on a MIMO system greatly hindering efforts to reduce conservatism if information about the inner structure of Δ is available.

The second uncertainty class, element-by-element structured uncertainties, provides a rich source of information about Δ. This class is defined as

$$D_s = \{\Delta = \{\delta_{ij}\} \mid |\delta_{ij}| \leq p_{ij} \in \Re^+\}$$

where a magnitude bound is placed on each element of uncertainty matrix Δ allowing information on the inner structure of Δ to remain a part of the stability analysis process. Class D_s provides a more realistic representation of real world uncertainty than that of the unstructured uncertainty class.

To illustrate the advantages of structured uncertainties over unstructured uncertainties, consider some process plant with two independent control valves whose flow rate is known within $\pm 10\%$ [21]. As a structured uncertainty, Δ could be correctly represented as

$$\Delta = \begin{bmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{bmatrix}, \quad |\delta_i| \leq 0.1.$$

Placing a norm bound of $\|\Delta\|_2 \leq .1$ as required for unstructured uncertainties allows a number of impossible structures like

$$\Delta = \begin{bmatrix} 0 & 0 \\ 0.1 & 0 \end{bmatrix}, \quad \text{or} \quad \Delta = \frac{1}{2} \cdot \begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.1 \end{bmatrix}.$$

Since there are two uncertainties, not one or four, these do not correspond to any realizable uncertainty configuration so a stability analysis accounting for such impossible structures would be unnecessarily conservative.

The third uncertainty class, block structured uncertainties, is actually a superset of the first two classes. Mathematically represented as

$$D_b = \{\Delta \mid \|\Delta_{ij}\|_2 \leq \delta_{ij} \in \Re^+\}$$

the class D_b may have unstructured submatrices, Δ_{ij}, with arbitrary dimension. For the case where $size(\Delta_{ij}) = 1 \times 1$, this class reduces simply to D_s. Likewise, for $size(\Delta_{ij}) = n \times n$ where n is the size of Δ, class D_b reduces to class D_u. Block structured uncertainties advanced from work done by Doyle [15] and they have become an important area of study due to their quite general nature.

Most real-world system uncertainties may be cast in one of the three classes discussed above. However, choosing the particular class for a specific problem is seldom straightforward and assigning actual magnitudes for the elements of the chosen structure can be more difficult still. Comparing the behavior of nominal plant G_o with an uncertainty matrix Δ to that of the actual system provides insight into the accuracy of the chosen model although a large number of candidate plants and uncertainty structures may satisfactorily match experimental data gathered during the identifi-

cation process. Narrowing down which of these candidate model combinations best approximates the actual system becomes a process of model invalidation as those models that fail to satisfy some matching criteria are eliminated. No discussion of how a designer should choose one model over another will be presented here since this is a complete field unto itself. However, a description of the problem by Smith and Doyle may be found in [22].

While most real-world systems may be characterized by one or more of the uncertainty classes described here, it must be noted that other models offering improved model agreement exist for some specific problems. Most notable of these other models are those that deal with combinations of both real and complex uncertainties and those that address repeated or dependent uncertainty structures. In the three uncertainty structures discussed above, unstructured, element-by-element structured, and general block structured, some form of bound was placed on the magnitude of the uncertainty. Such magnitude bounds allow the phases of the uncertainty elements or blocks to vary freely between 0 and 2π so that worst case situations can be addressed. For many uncertainty sources like unmodelled dynamics, this phase freedom is mandatory to establish necessary and sufficient stability conditions. However, if some or all elements of an uncertainty structure are known to be strictly real, this allowable phase variation may provide for an overly conservative stability analysis [23].

A method for treating strictly real uncertainties was proposed by De Gaston and Safonov [24]. This method transforms the perturbed plant G_p into a series of convex hulls in the complex plane and, through an iterative process, determines a stabil-

ity margin that indicates the largest scalar multiplier of Δ for which stability is guaranteed. This approach provides necessary and sufficient stability conditions but only if an infinite number of iterations are performed. Fortunately, usable results are generally obtained for a finite number of iterations though the method is still computationally intensive.

For uncertainties consisting of combinations of real and complex elements, alternative formulations of the structured singular value have been proposed by several authors [23, 25, 26]. While these approaches appear to reduce the conservatism over methods that consider only complex uncertainty structures, a number of theoretical and computational challenges must be resolved before they can be reliably applied to engineering problems.

In addition to special uncertainty models addressing real elements, the case of repeated or dependent complex uncertainty models requires special consideration. This class is actually a subset of the other structured uncertainty classes with the distinction that two or more of the elements or blocks of Δ are constrained such that their phases must always be the same [27]. Thus, while the phases may vary between 0 and 2π, they may not do so independently of one another.

While the real and repeated uncertainty element models provide an improved stability analysis for their particular cases, the two structured uncertainty classes presented earlier account for a wide range of important problems. Therefore, only their treatment will be considered in the remainder of this work.

CHAPTER 3
SINGULAR VALUE TECHNIQUES

3.1 The Singular Value Decomposition

As discussed in Section 2.3, the necessary and sufficient stability criterion

$$\sup_{\Delta} \rho(M\Delta) < 1 \tag{3.1}$$

is nonconvex making it difficult if not impossible to determine the range of permissible Δs. Rather than giving up on this approach for uncertain multivariable systems, it is possible to continue the analysis using relationships involving the singular value decomposition. These relationships start out as conservative upper bounds but the results of Chapter 4 show that this conservatism can generally be eliminated completely. The singular value decomposition is defined for any matrix $A \in \mathcal{C}^{n\times n}$ as

$$A = X\Sigma Y^H \tag{3.2}$$

where X and Y are unitary matrices the columns of which form the left and right singular vectors of A respectively. The diagonal matrix Σ contains the singular values of A in decreasing order of magnitude. The singular values and vectors of A can be determined in terms of the eigenvalues and eigenvectors of the hermetian forms A^HA and AA^H as

$$\begin{aligned} A^HAY_i &= \sigma_i^2 Y_i \\ AA^HX_i &= \sigma_i^2 X_i. \end{aligned} \tag{3.3}$$

Other useful relationships of the singular values and corresponding singular vectors include:

$$\begin{aligned} AY_i &= \sigma_i X_i \\ Y_i^H A^H &= \sigma_i X_i^H \\ A^H X_i &= \sigma_i Y_i \\ X_i^H A &= \sigma_i Y_i^H. \end{aligned} \tag{3.4}$$

Definition 3.1 The 2-norm of $A \in C^{n\times n}$, $\|A\|_2$, is defined as

$$\|A\|_2 = \max_{0\neq x\in C^n} \frac{\|Ax\|_2}{\|x\|_2} \tag{3.5}$$

where $\|x\|_2$ is the vector 2-norm defined as

$$\|x\|_2 = \sqrt{|x_1|^2 + \cdots + |x_n|^2}.$$

Also, from Definition 3.1 comes the important inequality equation

$$\|Ax\|_2 \leq \|A\|_2\|x\|_2. \tag{3.6}$$

The following theorem relates the 2-norm of a matrix to its maximum singular value.

Theorem 3.1: For any matrix $A \in C^{n\times n}$

$$\overline{\sigma}(A) = \|A\|_2.$$

Proof: From Equation 3.2, matrix A may be written as $A = X\Sigma Y^H$. The 2-norm of a matrix is invariant to unitary transformations so

$$\|A\|_2 = \|X\Sigma Y^H\|_2 = \|\Sigma\|_2.$$

From Definition 3.1, this may be written as

$$\|A\|_2 = \|\Sigma\|_2 = \max_{0 \neq x \in C^n} \frac{\|\Sigma x\|_2}{\|x\|_2} = \max_{0 \neq x \in C^n} \frac{\sqrt{\sigma_1^2 |x_1|^2 + \cdots + \sigma_n^2 |x_n|^2}}{\sqrt{|x_1|^2 + \cdots + |x_n|^2}}.$$

Choosing $x = e_1$, the first standard basis vector gives $\frac{\|\Sigma x\|_2}{\|x\|_2} = \sigma_1(A)$ while any other choice for x gives $\frac{\|\Sigma x\|_2}{\|x\|_2} \leq \sigma_1(A)$. Therefore, $\|A\|_2 = \sigma_1(A) = \overline{\sigma}(A)$. ■

Using this result, the following lemma relates $\overline{\sigma}(A)$ to $\rho(A)$.

Lemma 3.1 For any matrices A, $B \in C^{n \times n}$,

$$\rho(AB) \leq \overline{\sigma}(AB). \tag{3.7}$$

Proof: From the eigenvalue/eigenvector equation we have

$$ABW_i = \lambda_i W_i.$$

where λ_i is an eigenvalue of (AB) and W_i the corresponding eigenvector. Taking the 2-norm of both sides gives

$$\|ABW_i\|_2 = \|\lambda_i W_i\|_2 = |\lambda_i| \|W_i\|_2.$$

Equation 3.6 allows this to be written as an inequality

$$\|AB\|_2 \|W_i\|_2 \geq \|ABW_i\|_2 = |\lambda_i| \|W_i\|_2.$$

Applying Theorem 3.1 along with the definition of the spectral radius gives

$$\overline{\sigma}(AB) \geq \rho(AB). \quad ■$$

Finally, the 2-norm of a matrix product is related to the product of the 2-norms as shown in the following lemma.

Lemma 3.2 For $A, B \in C^{n\times n}$

$$\overline{\sigma}(AB) \leq \overline{\sigma}(A)\overline{\sigma}(B). \tag{3.8}$$

Proof: Again, from Definition 3.1

$$\|AB\|_2 = \max_{0\neq x\in C^n} \frac{\|ABx\|_2}{\|x\|_2} \leq \max_{0\neq x\in C^n} \frac{\|A\|_2\|Bx\|_2}{\|x\|_2} = \|A\|_2 \max_{0\neq x\in C^n} \frac{\|Bx\|_2}{\|x\|_2}$$

or

$$\|AB\|_2 \leq \|A\|_2\|B\|_2. \quad \blacksquare$$

Replacing A and B with M and Δ respectively in Equations 3.7 and 3.8, the following expression must hold:

$$\rho(M\Delta) \leq \overline{\sigma}(M\Delta) \leq \overline{\sigma}(M)\overline{\sigma}(\Delta).$$

From these relationships, the necessary and sufficient stability criterion of Equation 3.1 can be written as the following sufficient stability conditions

$$\sup_{\Delta} \overline{\sigma}(M)\overline{\sigma}(\Delta) < 1 \tag{3.9}$$

and

$$\sup_{\Delta} \overline{\sigma}(M\Delta) < 1. \tag{3.10}$$

Since Equations 3.9 and 3.10 only guarantee sufficient stability conditions for structured uncertainties, additional manipulations are required to reduce conservatism with the intent of regaining the necessary stability condition. Two techniques that address this problem are nonsimilarity scaling introduced by Kouvaritakis and Latchman [16] and similarity scaling advanced by Doyle [15]. Both techniques rely on the fact that while the eigenvalues (and hence the spectral radius) of a matrix are unaffected by similarity transformations, the singular values are not necessarily

preserved and may, in fact, be reduced. The nonsimilarity scaling technique will be discussed first with the analysis limited to element by element bounded uncertainties.

3.2 Structured Uncertainties with Nonsimilarity Scaling

The nonsimilarity scaling formulation starts with the necessary and sufficient stability criterion of Equation 3.1. Introducing positive, diagonal matrices R and L to this expression gives the equivalent stability condition

$$\sup_{\Delta} \rho(M\Delta) = \sup_{\Delta} \rho(R^{-1}ML^{-1}L\Delta R) < 1.$$

This may be rewritten in the form of Equation 3.9 giving the sufficient stability criterion

$$\sup_{\Delta} \overline{\sigma}(R^{-1}ML^{-1})\overline{\sigma}(L\Delta R) < 1. \tag{3.11}$$

Equation 3.11 no longer conforms to the similarity transformation structure: hence the name "nonsimilarity scaling." However, the free elements of R and L still allow the conservatism gap between the maximum singular value upper bound and the spectral radius lower bound to be reduced. Also, the choice of positive values for the elements of R and L allows a simplification based on the following lemmas and Theorem 3.2.

Lemma 3.3: For all $A \in \mathcal{C}^{n\times n}$ and $B \in \Re^{n\times n}$ where $b_{ij} \geq 0$ and $|a_{ij}| \leq b_{ij}$ for all i, j, with x_+ defined as $(|x_1|, \cdots, |x_n|)^T$,

$$\|Ax\|_2 \leq \|Bx_+\|_2 \text{ for all } x \in \mathcal{C}^n. \tag{3.12}$$

Proof: From the definition of a vector 2-norm

$$\|Ax\|_2 = \sqrt{|a_{11}x_1 + \cdots + a_{1n}x_n|^2 + \cdots + |a_{n1}x_1 + \cdots + a_{nn}x_n|^2}$$

$$\|Bx_+\|_2 = \sqrt{|\ [b_{11}|x_1| + \cdots + b_{1n}|x_n|]\ |^2 + \cdots + |\ [b_{n1}|x_1| + \cdots + b_{nn}|x_n|]\ |^2}.$$

The second equation contains only nonnegative numbers so the sum of the individual terms cannot decrease. The first equation, however, contains elements that may be negative so the sum of the individual terms may decrease. Since the elements of B are greater than or equal to the corresponding elements of A, the lemma must hold. ■

Lemma 3.4 Using the definition of x_+ from Lemma 3.3,

$$\|x\|_2 = \|x_+\|_2 \text{ for all } x \in C^n. \tag{3.13}$$

Proof: Again, from the definition of the vector 2-norm

$$\|x\|_2 = \sqrt{|x_1|^2 + \cdots + |x_n|^2} = \|x_+\|_2. \quad ■$$

Lemma 3.5 For A and B defined as in Lemma 3.3,

$$\frac{\|Ax\|_2}{\|x\|_2} \leq \|B\|_2 \text{ for all } 0 \neq x \in C^n. \tag{3.14}$$

Proof: From Definition 3.1,

$$\|B\|_2 = \max_{0 \neq x \in C^n} \frac{\|Bx\|_2}{\|x\|_2}.$$

Since all elements of B are nonnegative, this may be combined with Lemmas 3.3 and 3.4 to give

$$\|B\|_2 = \max_{0 \neq x_+ \in C^n} \frac{\|Bx_+\|_2}{\|x_+\|_2} \geq \frac{\|Ax\|_2}{\|x\|_2}. \quad ■$$

Theorem 3.2: Given $\Delta \in C^{n\times n}$, positive, diagonal matrices $L, R \in R^{n\times n}$ and $P \in \Re^{n\times n}$ where $p_{ij} \geq 0$ and $|\delta_{ij}| \leq p_{ij}$ for all i, j,

$$\overline{\sigma}(L\Delta R) \leq \overline{\sigma}(LPR). \tag{3.15}$$

Proof: This proof follows directly from Lemmas 3.3, 3.4, and 3.5 with A replaced by $L\Delta R$ and B replaced by LPR so that

$$\|L\Delta R\|_2 = \max_{0\neq x\in C^n} \frac{\|\Delta x\|_2}{\|x\|_2} \leq \max_{0\neq x_+\in C^n} \frac{\|LPRx_+\|_2}{\|x_+\|_2} = \|LPR\|_2. \quad \blacksquare$$

Theorem 3.2 and the definition of element-by-element structured uncertainties from Section 2.4 allow the direct substitution of P for Δ in Equation 3.11.

Substituting P for Δ would appear to add additional conservatism to Equation 3.11. However, Lemma 3.6 shows that this is not the case.

Lemma 3.6 Given $L\Delta R$ and LPR defined as in Theorem 3.2, the following equality must hold:

$$\sup_{\Delta} \overline{\sigma}(L\Delta R) = \overline{\sigma}(LPR). \tag{3.16}$$

Proof: The freedom of the phases of each element of Δ allows $\delta_{ij} = |\delta_{ij}| = p_{ij}$ for all i, j. Therefore, by individually adjusting the phases of each element of Δ, the equality of Equation 3.16 is established. $\blacksquare$

Since P always represents a realizable variation of Δ, there is actually no added conservatism by substituting P for Δ. Therefore the sufficient stability condition of Equation 3.11 becomes

$$\inf_{L,R} \overline{\sigma}(R^{-1}ML^{-1})\overline{\sigma}(LPR) < 1 \tag{3.17}$$

over all design frequencies. This infimization over L and R requires only $2(n-1)$ optimization variables because one diagonal element of both L and R may be held constant without loss of generality. For the case of distinct $\overline{\sigma}(R^{-1}ML^{-1})$, Equation 3.17 has been shown to converge to the necessary and sufficient condition of Equation 3.1. This optimal condition is characterized by the major principal direction alignment (MPDA) theorem proposed by Kouvaritakis and Latchman [17]. Due to the importance of this theorem it will be discussed in detail in Chapter 4.

3.3 Structured Uncertainties with Similarity Scaling

The previous section outlined the nonsimilarity scaling technique which applies to systems with element-by-element structured uncertainties. A second approach to solving Equation 3.1 through singular value techniques involves diagonal similarity scaling where only one scaling matrix is used. Advanced by Doyle [15], this technique requires that the uncertainty matrix Δ be diagonal with norm bounds on the diagonal elements or blocks. An important advantage of diagonal similarity scaling is its ability to address the general block structured uncertainties discussed in Section 2.4. While the requirement for diagonalized uncertainties appears to severely restrict the applicability of this technique, it is possible to diagonalize any uncertainty matrix through the use of simple, eigenvalue preserving transformations. Properties of these transformations will be discussed in Section 3.5.

For diagonal uncertainty matrix Δ_d, diagonal scaling matrix D may be introduced allowing the necessary and sufficient stability conditions of Equation 3.1 to be written

as

$$\sup_{\Delta_d} \rho(M\Delta_d) = \sup_{\Delta_d} \rho(DM\Delta_d D^{-1}) < 1 \tag{3.18}$$

and the sufficient condition of Equation 3.10 as

$$\sup_{\Delta_d} \overline{\sigma}(DM\Delta_d D^{-1}) < 1. \tag{3.19}$$

Taking advantage of the diagonal nature of Δ_d permits a decomposition of the form

$$\Delta_d = P_d U_d$$

where $p_{d_{ii}} \geq |\delta_{d_{ii}}|$ for all i and U_d is a diagonal unitary matrix. This allows P_d to retain the magnitude information and U_d the phase information of Δ_d. Substituting $P_d U_d$ for Δ_d in Equation 3.19 gives the sufficient stability condition

$$\sup_{\Delta_d} \rho(M\Delta_d) \leq \sup_{U_d} \overline{\sigma}(DMP_d U_d D^{-1}) < 1.$$

Noting that diagonal matrices commute with each other, the positions of U_d and D^{-1} may be exchanged making the condition

$$\sup_{\Delta_d} \rho(M\Delta_d) \leq \sup_{U_d} \overline{\sigma}(DMP_d D^{-1} U_d) < 1. \tag{3.20}$$

However, the 2-norm, and hence the maximum singular value, of a matrix is invariant to unitary transformations removing the dependence on U_d. Combining M and P_d into $M_a = MP_d$, the condition of Equation 3.20 becomes

$$\sup_{\Delta_d} \rho(M\Delta_d) \leq \inf_{D} \overline{\sigma}(DM_a D^{-1}) < 1. \tag{3.21}$$

Since the spectral radius of $M_a U_d$ is always bounded from above by the maximum singular value of $DM_a D^{-1}$, the free elements of D may be used to reduce the conservatism gap between the maximum singular value upper bound and the spectral radius

lower bound of Equation 3.21. For the general case where a stationary point occurs at the "inf," the resulting optimal value of $\overline{\sigma}(DM_aD^{-1})$ is known as the "structured singular value" of M_a (denoted $\mu(M_a)$).

The original definition of the structured singular value had the form [15]

$$\mu(M_a) = \left\{ \begin{array}{ll} 0 & \text{if no } \Delta \text{ solves } det\{[I + M\Delta]\} = 0 \\ \text{else} & (\min_{\Delta} \{\overline{\sigma}(\Delta) \mid det\{[I + M\Delta]\} = 0\})^{-1} \end{array} \right\}.$$

While this expression is rather unwieldy, a more useful alternative formulation is given by

$$\mu(M_a) = \sup_{U_d} \rho(M_aU_d) \leq \inf_{D} \overline{\sigma}(DM_aD^{-1}). \tag{3.22}$$

The right-hand side of Equation 3.22 has been proven convex (with respect to scaling matrix e^D) [28, 29] and the MPDA formulation for similarity scaling shows that the inequality becomes an equality at the "inf" for stationary points of $\overline{\sigma}(DM_aD^{-1})$ [17].

Chapter 4 reviews conditions required for the optimal solution to Equation 3.22. However, an interesting suboptimal solution was proposed by Safonov involving the Frobenius norm minimization of DM_aD^{-1} [14]. Since the Frobenius norm of a matrix provides an upper bound on the 2–norm (and hence the maximum singular value) of the matrix, reducing the Frobenius norm of DM_aD^{-1} generally serves to reduce $\overline{\sigma}(DM_aD^{-1})$ as well. An iterative algorithm introduced by Osborne performs such a reduction by choosing elements of D that equalize the row and column sums of DM_aD^{-1} [30]. While this procedure produces a conservative upper bound for $\mu(M_a)$, it is numerically inexpensive since no singular value decompositions are required.

3.4 Block Structured Uncertainties with Similarity Scaling

As discussed in Section 2.4, the family of block structured uncertainties represents the most general class since it is a superset of both the unstructured and element-by-element structured uncertainty classes. Let Δ_b represent a block diagonal matrix containing both scalar and nonscalar uncertainty elements. As in Section 3.3, this can be decomposed giving

$$\Delta_b = P_b U_b$$

where P_b contains the magnitude bounds and U_b the phase information of Δ_b. An important difference between this decomposition and that of Section 3.3 is the requirement for P_b and U_b to reflect the block structure of Δ_b. For this configuration, the diagonal scaling matrix D_b must also reflect the block structure requiring repeated elements of D_b as necessary to correspond to the size of the individual blocks in Δ_b. With the exception of maintaining the block structure, this formulation is identical to that for element-by-element uncertainties and, in fact, the block form of Equation 3.22 simply becomes

$$\mu(M_b) = \sup_{U_b} \rho(M_b U_b) \leq \inf_{D_b} \overline{\sigma}(D_b M_b D_b^{-1}) \tag{3.23}$$

where $M_b = M P_b$.

3.5 Diagonalizing Transformations

As mentioned earlier, it is always possible to reconfigure any structured Δ into diagonalized form. First consider uncertainty class D_s described in Section 2.4 where $\Delta \in C^{n \times n}$ contains n^2 complex elements δ_1 to δ_{n^2}. Next define $\Delta_d \in C^{n^2 \times n^2} =$

$diag[\delta_1, \ldots, \delta_{n^2}]$. It is always possible to relate Δ_d to Δ using transformation matrices $E_1 \in \Re^{n \times n^2}$ and $E_2 \in \Re^{n^2 \times n}$ containing only $1's$ and $0's$ such that $E_1 \Delta_d E_2 = \Delta$. From this relationship, the necessary and sufficient stability conditions of Equation 3.1 may be written as

$$\sup_{\Delta_d} \rho(M E_1 \Delta_d E_2) < 1.$$

At this point, some means of recovering the simplifying properties of diagonal matrices required by Equation 3.20 is needed. The following theorem relates the eigenvalues (and hence the spectral radius) of matrix products (AB) and (BA).

Theorem 3.3: Given $A \in C^{n \times m}$ and $B \in C^{m \times n}$,

$$\lambda_i(AB) = \lambda_i(BA) \quad i = 1, \ldots, n.$$

Proof: Define λ_i and $\hat{\lambda}_j$ as the eigenvalues of (AB) and (BA) respectively. Then

$$(AB)W_i = \lambda_i W_i \quad i = 1, \ldots, n$$

$$(BA)Y_j = \hat{\lambda}_j Y_j \quad j = 1, \ldots, m$$

where $W_i \in C^{n \times 1}$ and $Y_j \in C^{m \times 1}$ form the corresponding eigenvectors. Multiply the first equation by B on the left giving

$$(BAB)W_i = \lambda_i B W_i.$$

Vector $BW_i \in C^{m \times 1}$ must be an eigenvector corresponding to the ith eigenvalue of (BA) as well as the ith eigenvalue of (AB). Since the eigenvalues of a matrix are unique, $\lambda_i(AB) = \hat{\lambda}_i(BA)$ for $i = 1, \ldots, n$. ■

Using this result, replace A by ME_1P_d and B by E_2 so the spectral radius equations become

$$\rho(M\Delta) = \rho(E_2ME_1\Delta_d).$$

Next, define $M_a = E_2ME_1$ and the necessary and sufficient stability condition of Equation 3.1 may be written

$$\sup_{\Delta_d} \rho(M_a\Delta_d) < 1.$$

Starting at this point, the analysis of Section 3.3 can then continue as before.

While this diagonalizing transformation is straightforward, it has the unfortunate consequence of increasing the system size from $n \times n$ to as much as $n^2 \times n^2$ with a resultant increase in computational requirements to compute the maximum singular values. Also, the number of optimization variables required to infimize Equation 3.22 can increase to $n^2 - 1$ versus $2(n-1)$ for nonsimilarity scaling. (As with matrices L and R of Equation 3.17, a single element of D may be arbitrarily fixed allowing the elimination of one optimization variable). Even with these disadvantages, the ability to manage block structured uncertainties makes similarity scaling an invaluable tool in the analysis of uncertain systems.

CHAPTER 4
THE MAJOR PRINCIPAL DIRECTION ALIGNMENT PROPERTY

4.1 Application to Similarity Scaling

The use of singular value techniques in the form of nonsimilarity and similarity scaling has the disadvantage of initially reducing the necessary and sufficient stability conditions of Equation 3.1 to simply sufficient conditions. For convenience, the relevant equations are repeated.

$$\sup_{\Delta} \rho(M\Delta) \leq \inf_{L,R} \overline{\sigma}(R^{-1}ML^{-1})\overline{\sigma}(LPR) \quad \text{Nonsimilarity Scaling} \tag{4.1}$$

$$\sup_{\Delta} \rho(M\Delta) \leq \inf_{D} \overline{\sigma}(DM_aD^{-1}) \quad \text{Similarity Scaling.} \tag{4.2}$$

Eliminating conservatism in these singular value formulations requires establishing conditions for which the inequalities hold with equality. This section reveals that achieving equality in Equations 4.1 and 4.2 is simply a matter of invoking the MPDA property [17]. Although MPDA conditions have been shown to exist for both nonsimilarity and similarity formulations, this discussion will focus only the similarity scaling techniques.

Definition 4.1: The major right principal direction, Y_1, and the major left principal direction, X_1, of matrix A are the right and left singular vectors of matrix A corresponding to $\overline{\sigma}(A)$.

Theorem 4.1: For any matrix $A \in \mathcal{C}^{n \times n}$

$$\rho(A) = \overline{\sigma}(A)$$

if and only if X_1 and Y_1 are aligned within a scaling factor $e^{j\theta}$ such that

$$Y_1 = e^{j\theta} X_1. \tag{4.3}$$

Proof: For notational simplicity, denote $\overline{\sigma}(A)$ by $\overline{\sigma}$ and $\rho(A)$ by ρ. Vectors X_1 and Y_1 from the singular value decomposition must be unique with respect to one another within a scaling factor $e^{j\theta}$. Multiplying both sides of Equation 4.3 on the left by A gives

$$AY_1 = e^{j\theta} AX_1.$$

Also, the relationships of Section 3.1 show that $AY_1 = \overline{\sigma} X_1$ so this can be rewritten as

$$AX_1 = \overline{\sigma} e^{-j\theta} X_1$$

indicating that $\overline{\sigma} e^{-j\theta}$ actually corresponds to an eigenvalue of A. Since eigenvalues of a matrix are always magnitude bounded by the maximum singular value of the matrix, sufficiency of the theorem is established.

Establishing the necessity of the theorem starts with the assumption that $\overline{\sigma} = \rho$ so some eigenvector Z_1 exists where

$$AZ_1 = e^{j\phi} \overline{\sigma} Z_1 \tag{4.4}$$

and

$$Z_1^H A^H = e^{-j\phi} \overline{\sigma} Z_1^H. \tag{4.5}$$

Multiplying Equation 4.4 on the left by Equation 4.5 and noting the $A^H A = Y\Sigma^2 Y$ provides the following expression:

$$\frac{Z_1^H A^H A Z_1}{Z_1^H Z_1} = \frac{Z_1^H Y \Sigma^2 Y^H Z_1}{Z_1^H Y Y^H Z_1} = \overline{\sigma}^2.$$

This may be simplified by defining a new vector $W_1 = Y_1^H Z_1$ to give

$$\frac{W_1^H \Sigma^2 W_1}{W_1^H W_1} = \overline{\sigma}^2. \tag{4.6}$$

Equation 4.6 can only be satisfied if $W_1 = e^{j\psi} e_1$ where e_1 is the first standard basis vector. Therefore, $Z_1 = YW_1 = e^{j\psi} Y e_1 = e^{j\psi} Y_1$. Substituting back into Equation 4.4 then gives

$$Ae^{j\psi} Y_1 = \overline{\sigma} e^{j\psi} X_1 = e^{j\phi} \overline{\sigma} e^{j\psi} Y_1$$

or

$$X_1 = e^{j\psi} Y_1$$

establishing the necessity of the theorem. ■

Corollary 4.1: For the case of repeated singular values of $A \in C^{n \times n}$ where q denotes the multiplicity of σ_1,

$$\rho(A) = \sigma_1(A), \ldots, \sigma_q(A)$$

if and only if the subspaces spanned by $X_1, \ldots, X_q$ and $Y_1, \ldots, Y_q$ are aligned within a scaling factor $e^{j\theta}$.

Proof: From Equation 3.2 the singular vectors of matrix A with a repeated maxi-

mum singular value are defined as

$$A^H A Y_1 = \sigma_1^2 Y_1 \qquad AA^H X_1 = \sigma_1^2 X_1$$
$$\vdots \qquad \text{and} \qquad \vdots \quad .$$
$$A^H A Y_q = \sigma_1^2 Y_q \qquad AA^H X_q = \sigma_1^2 X_q$$

Therefore, linear combinations of the first q columns of X or Y form legitimate solutions for X_1 and Y_1 respectively. Combining these solutions with Theorem 4.1 completes the proof. ■

The MPDA property therefore provides an analytical test for equality in Equation 3.21 through which the necessary and sufficient stability conditions of Equation 3.1 apply. This allows the infimization of Equation 3.22 to be performed while providing a direct test as to whether the structured singular value, $(\mu(M_a))$, has been achieved. It would be desirable if this infimization was somehow guaranteed to invoke MPDA and hence attain $\mu(M_a)$. Fortunately, such a guarantee is provided by the following theorem.

Theorem 4.2: The infimizing solution with respect to diagonal scaling matrix D of stability criterion

$$\inf_D \overline{\sigma}(DM_aD^{-1}) < 1 \tag{4.7}$$

is both necessary and sufficient provided that $\frac{\partial \overline{\sigma}}{\partial d_i} = 0$ at the "inf."

Proof: The sufficiency of the theorem was shown in Equation 3.21. Necessity is established as long as MPDA can be achieved. For simple $\overline{\sigma}(DM_aD^{-1})$, the optimal solution of Equation 4.7 occurs when

$$\frac{\partial \overline{\sigma}^2}{\partial d_i}(DM_aD^{-1}) = 0$$

indicating a stationary point. As mentioned in Section 3.3, the left-hand side of Equation 4.7 is convex with respect to scaling matrix e^D so any local minimum is the global minimum. Therefore the only stationary point must correspond to the optimal solution. By direct differentiation, the stationarity condition becomes

$$\frac{\partial \overline{\sigma}^2}{\partial d_i}(\tilde{M}_a) = \frac{2Y_1^H}{d_i}\left[\tilde{M}_a^H E_i \tilde{M}_a - \overline{\sigma}^2 E_i\right] Y_1 = 0 \tag{4.8}$$

where $\tilde{M}_a = DM_aD^{-1}$, and E_i is a square matrix containing a "1" in the iith position and "0s" elsewhere. (Note: the actual differentiation is omitted here but the complete differentiation of a similar function appears in Section 5.5).

Examining conditions at which Equation 4.8 equals zero reveals the requirement that

$$|x_{1i}| = |y_{1i}| \quad i = 1, \ldots, n \tag{4.9}$$

which is in fact the MPDA requirement that the magnitudes of the elements of X_1 and Y_1 be equal.

Completion of the proof from this point requires that some U_d exists such that the phase alignment of X_1 and Y_1 occurs. The existence of such a U_d then guarantees that $\rho(M_aU_d) = \overline{\sigma}(DM_aD^{-1}U_d) = \mu(M_a)$. From the singular value decomposition, the principal singular vectors are related as

$$\begin{aligned} (DM_aD^{-1})Y_1 &= \overline{\sigma}X_1 \\ X_1^H(DM_aD^{-1}) &= \overline{\sigma}Y_1^H. \end{aligned}$$

Introducing a diagonal unitary matrix, U_d, allows these equations to be rewritten as

$$(DM_aD^{-1}U_d)U_d^HY_1 = \overline{\sigma}X_1$$

$$X_1^H(DM_aD^{-1}U_d) = \overline{\sigma}Y_1^HU_d. \tag{4.10}$$

From Equation 4.10 it is apparent that the diagonal elements of U_d may be used to alter the phases of X_1 and Y_1 while continuing to satisfy the MPDA requirement that the magnitudes of the corresponding elements of X_1 and Y_1 be equal. The ability to achieve independent phase alignment of X_1 and Y_1 by U_d completes the proof. ■

The result of this theorem is that optimization of Equation 4.7 has the effect of directly invoking MPDA at the "inf" for simple $\overline{\sigma}$, guaranteeing the necessary and sufficient stability conditions. Cases with nonsimple $\overline{\sigma}$ (known as "cusps") do not necessarily contain a stationary point at the "inf" and must be analyzed using different techniques. A more complete discussion of cusping systems appears in Chapter 8.

4.2 Block Structured Uncertainties

By following the MPDA development for scalar uncertainties, a similar formulation results for general block structured uncertainties. The main difference is the requirement of maintaining the individual block structures rather than considering each element of Δ individually. As discussed in Section 3.4, this block structure must be reflected in the scaling matrices as well. The optimal solution is then characterized by the equalization of the 2-norms in the corresponding blocks of the left and right principal vectors X_1, and Y_1 respectively. The same guarantee of achieving equality

at the optimal solution of Equation 3.23 for simple $\overline{\sigma}(D_b M_b D_b^{-1})$ applies as in the element-by-element structured uncertainty case.

4.3 Examples

The following examples illustrate the methods described in the previous sections.

Example 4.1: Let randomly generated matrices $M \in \mathcal{C}^{3\times3}$ be chosen such that

$$M = \begin{bmatrix} .063 + .156\jmath & -.322 + .480\jmath & .585 + .526\jmath \\ .726 - .514\jmath & -.323 - .344\jmath & .150 - .469\jmath \\ .189 - .463\jmath & .053 - .577\jmath & -.236 - .056\jmath \end{bmatrix}.$$

Next chose a random element-by-element structured uncertainty matrix $\Delta \in \mathcal{C}^{3\times3}$ with corresponding $P \in \Re^{3\times3}$ written as

$$\begin{bmatrix} |\delta_1| & |\delta_2| & |\delta_3| \\ |\delta_4| & |\delta_5| & |\delta_6| \\ |\delta_7| & |\delta_8| & |\delta_9| \end{bmatrix} \leq \begin{bmatrix} 2.99 & 3.03 & 0.54 \\ 1.65 & 1.87 & 3.41 \\ 1.90 & 1.20 & 1.37 \end{bmatrix} = P.$$

Formulated as a nonsimilarity scaling problem,

$$\mu(M\Delta) \leq \inf_{L,R} \overline{\sigma}(R^{-1}ML^{-1})\overline{\sigma}(LPR)$$

the diagonal elements of the optimal R and L scaling matrices are found to be

$$R = \begin{bmatrix} r_1 = 1 \\ r_2 = 1.151 \\ r_3 = 1.174 \end{bmatrix}, \; L = \begin{bmatrix} l_1 = 1 \\ l_2 = 0.944 \\ l_3 = 1.203 \end{bmatrix}.$$

The maximum singular values do not repeat for this example so

$$\mu(M_a) = \overline{\sigma}(R^{-1}ML^{-1})\overline{\sigma}(LPR) = 8.25.$$

This same problem may be recast into a similarity scaling form by first selecting transformation matrices E_1 and E_2 as

$$E_1 = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Expanding M and P as described in Section 3.3, the optimal solution of Equation 3.22 requires scaling matrix D with diagonal elements

$$D = \begin{bmatrix} d_1 = 1 \\ d_2 = 0.906 \\ d_3 = 0.406 \\ d_4 = 0.719 \\ d_5 = 0.688 \\ d_6 = 0.988 \\ d_7 = 0.783 \\ d_8 = 0.560 \\ d_9 = 0.636 \end{bmatrix}.$$

As before, $\mu(M_a) = \overline{\sigma}(DM_aD^{-1}) = 8.25$. The absolute values of the left and right principal directions for this optimal solution are

$$|X_1| = \begin{bmatrix} 1 \\ 1.119 \\ 0.444 \\ 0.719 \\ 0.850 \\ 1.080 \\ 0.783 \\ 0.691 \\ 0.695 \end{bmatrix} = |Y_1| = \begin{bmatrix} 1 \\ 1.119 \\ 0.444 \\ 0.719 \\ 0.850 \\ 1.080 \\ 0.783 \\ 0.691 \\ 0.695 \end{bmatrix}.$$

As required by the MPDA theory, $|X_1| = |Y_1|$ at the "inf" guaranteeing that some U_d exists such that $\rho(M_a\hat{U}_d) = \overline{\sigma}(\hat{D}M_a\hat{D}^{-1}) = \mu(M_a)$.

Example 4.2: Using the M matrix from Example 4.1, define a block structured uncertainty matrix as

$$\Delta = \begin{bmatrix} \delta_1 & \delta_2 & \delta_3 \\ \delta_4 & & \\ \delta_5 & & \Delta_b \end{bmatrix}$$

where δ_1 through δ_5 are scalar uncertainties such that $|\delta_i| \leq p_i$ and Δ_b is a 2×2 block with $\overline{\sigma}(\Delta_b) \leq p_b$. Block Δ_b may be represented as

$$\Delta_b = P_b U_b$$

where P_b is a 2×2 matrix with $\overline{\sigma}(P_b) = p_b$ and U_b is a 2×2 unitary matrix containing the phase information of Δ_b. We may now define a matrix P_Δ containing randomly generated bounds on the δ_i and Δ_b elements as

$$P_\Delta = \begin{bmatrix} 2.99 & 3.03 & 0.54 \\ 1.65 & & \\ 1.90 & & P_b \end{bmatrix}$$

with $\overline{\sigma}(P_b) = 1.87$. It is easily shown that the worst case structure for P_b in terms of $\mu(M_b)$ occurs when

$$P_b = \begin{bmatrix} 1.87 & 0 \\ 0 & 1.87 \end{bmatrix}$$

allowing a P matrix for this case to be written as

$$P = \begin{bmatrix} 2.99 & 3.03 & 0.54 \\ 1.65 & 1.87 & 0 \\ 1.90 & 0 & 1.87 \end{bmatrix}.$$

The 2×2 block structure constrains the last two terms of scaling matrix D_b to be identical thus eliminating one degree of freedom from the optimization. Also, the two zero elements of P allow the diagonalized matrix P_d (and hence M_b) to be of size 7×7 rather than 9×9.

Choose

$$E_1 = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

corresponding to $P_d = diag[2.99,\ 3.03,\ 0.54,\ 1.65,\ 1.9,\ 1.87,\ 1.87]$ so that $P = E_1 P_d E_2$. Expanded matrix M_b may then be formed as $M_b = E_2 M E_1 P_d$. Following the infimization of $\overline{\sigma}(D_b M_b D_b^{-1})$, the optimizing D_b and the absolute values of the elements of X_1 and Y_1 are found to be

$$D_b = \begin{bmatrix} d_1 = 1 \\ d_2 = 0.82 \\ d_3 = 0.41 \\ d_4 = 0.63 \\ d_5 = 0.87 \\ d_6 = 0.69 \\ d_7 = 0.69 \end{bmatrix}, \quad |X_1| = \begin{bmatrix} 0.427 \\ 0.527 \\ 0.187 \\ 0.269 \\ 0.373 \\ 0.442 \\ 0.313 \end{bmatrix}, \quad |Y_1| = \begin{bmatrix} 0.427 \\ 0.527 \\ 0.187 \\ 0.269 \\ 0.373 \\ 0.279 \\ 0.465 \end{bmatrix}$$

with $\mu(M_b) = 6.5$. Note that the first five terms of $|X_1|$ and $|Y_1|$ are equal as in the case of scalar uncertainty blocks under MPDA conditions. However, the presence of the 2×2 block causes the last two terms of $|X_1|$ and $|Y_1|$ to be related in the following manner:

$$|x_{61}|^2 + |x_{71}|^2 = |y_{61}|^2 + |y_{71}|^2 = 0.2937.$$

As described in Section 4.2, this equality between norms of the corresponding blocks of X_1 and Y_1 occurs whenever MPDA is established with block structured uncertainties guaranteeing the existence of some optimal $\hat{U}_b$ such that $\rho(M_b \hat{U}_b) = \overline{\sigma}(\hat{D}_b M_b \hat{D}_b^{-1}) = \mu(M_b)$ [31].

The material presented so far provides a motivation for computing the structured

singular value using singular value scaling techniques. However, even with the examples, very little insight can be gained into the computational issues involved with determining μ. In fact, although the structured singular value concept is currently being applied to the design of large engineering problems [23], a number of difficulties continue to hinder its application. The next chapter shows that knowledge of the uncertainty structure can lead to a reduction in the number of optimization variables required to compute μ using similarity scaling.

CHAPTER 5
REDUCTION IN THE NUMBER OF OPTIMIZATION VARIABLES REQUIRED TO COMPUTE THE STRUCTURED SINGULAR VALUE

5.1 Reduction of Optimization Variables

As mentioned earlier, similarity scaling techniques have the capability of handling general block structured uncertainties. While this advantage justifies the use of similarity scaling for systems with these uncertaint structures, the requirement that the uncertainty blocks be diagonalized results in a possibly substantial increase in overall system size. For scalar blocks this expansion transforms an $n \times n$ system to an $n^2 \times n^2$ system. Not only does this extended system size increase the floating point operation (flop) count for each singular value decomposition, but, as discussed in Section 3.5, it also increases the number of optimization variables from $2(n-1)$ for nonsimilarity scaling to $n^2 - 1$ for similarity scaling. Since M, Δ and hence $\mu(M_a)$ are frequency dependent functions, a frequency sweep of μ must be performed to determine the worst case condition. Depending on the operating frequency range of the control system under investigation, this frequency sweep may involve hundreds or thousands of individual points at which μ must be evaluated. Each of these points therefore represents a substantial cost in CPU time.

Since both nonsimilarity and similarity scaling methods produce identical results for element-by-element structured uncertainties, it would seem reasonable to conclude that some of the $n^2 - 1$ degrees of freedom required for similarity scaling are

actually redundant. Removing these unnecessary optimization variables, if possible, should provide for some corresponding reduction in flops required to compute μ using similarity scaling. The following theorem and subsequent proof show that, in fact, no more that $2(n-1)$ optimization variables are required to compute μ for both similarity and nonsimilarity scaling.

Theorem 5.1: Given M $\in C^{n\times n}$ and $\Delta \in C^{n\times n}$, where Δ is composed of 1×1 blocks, the solution for $\mu(M_a)$ in the optimization problem

$$\mu(M_a) \leq \inf_D \overline{\sigma}(DM_aD^{-1}) \tag{5.1}$$

can be found with no more than $2(n-1)$ free variables as long as a stationary point occurs at the "inf."

Proof: To reduce notational complexity this proof will be developed explicitly for the 2×2 case and the general $n \times n$ case will follow as a simple extension. Let

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}, \ \Delta = \begin{bmatrix} \delta_1 & \delta_2 \\ \delta_3 & \delta_4 \end{bmatrix}$$

$$\begin{bmatrix} |\delta_1| & |\delta_2| \\ |\delta_3| & |\delta_4| \end{bmatrix} \leq \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = P,$$

and $P_d = diag[p_{11}, p_{12}, p_{21}, p_{22}]$. For this P_d choose

$$E_1 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \ E_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Applying scaling matrix $D = diag[d_1, d_2, d_3, d_4]$ to M_a gives

$$DM_aD^{-1} = \begin{bmatrix} m_{11}p_{11} & \frac{d_1}{d_2}m_{11}p_{12} & \frac{d_1}{d_3}m_{12}p_{21} & \frac{d_1}{d_4}m_{12}p_{22} \\ \frac{d_2}{d_1}m_{21}p_{11} & m_{21}p_{12} & \frac{d_2}{d_3}m_{22}p_{21} & \frac{d_2}{d_4}m_{22}p_{22} \\ \frac{d_3}{d_1}m_{11}p_{11} & \frac{d_3}{d_2}m_{11}p_{12} & m_{12}p_{21} & \frac{d_3}{d_4}m_{12}p_{22} \\ \frac{d_4}{d_1}m_{21}p_{11} & \frac{d_4}{d_2}m_{21}p_{12} & \frac{d_4}{d_3}m_{22}p_{21} & m_{22}p_{22} \end{bmatrix}. \tag{5.2}$$

Also, from the singular value decomposition,

$$DM_aD^{-1} = X\Sigma Y^H \tag{5.3}$$

where the orthogonal matrices X and Y represent the left and right singular vectors of (DM_aD^{-1}) respectively and Σ is a positive diagonal matrix containing the singular values of (DM_aD^{-1}) in decreasing order of magnitude.

For the 2×2 case, matrix M must have rank≤ 2. Multiplying M on the left by E_2 and on the right by E_1P_d to form M_a therefore requires that rank$(M_a)\leq 2$ forcing $\sigma_3 = \sigma_4 = 0$. Expansion of Equation 5.3 into its individual elements gives

$$DM_aD^{-1} = \begin{bmatrix} x_{11}\overline{y}_{11}\sigma_1 + x_{12}\overline{y}_{12}\sigma_2 & \cdots & x_{11}\overline{y}_{41}\sigma_1 + x_{12}\overline{y}_{42}\sigma_2 \\ x_{21}\overline{y}_{11}\sigma_1 + x_{22}\overline{y}_{12}\sigma_2 & \cdots & x_{21}\overline{y}_{41}\sigma_1 + x_{22}\overline{y}_{42}\sigma_2 \\ x_{31}\overline{y}_{11}\sigma_1 + x_{32}\overline{y}_{12}\sigma_2 & \cdots & x_{31}\overline{y}_{41}\sigma_1 + x_{32}\overline{y}_{42}\sigma_2 \\ x_{41}\overline{y}_{11}\sigma_1 + x_{42}\overline{y}_{12}\sigma_2 & \cdots & x_{41}\overline{y}_{41}\sigma_1 + x_{42}\overline{y}_{42}\sigma_2 \end{bmatrix}. \tag{5.4}$$

Examination of Equation 5.2 shows that the ratio of the first and third rows is a scalar constant $\alpha_1 = d_1/d_3$. Since Equation 5.2 and Equation 5.4 both equal DM_aD^{-1}, the ratio of the first and third rows of Equation 5.4 must also be α_1. Equating the first row of Equation 5.4 to α_1 times the third row requires that

$$\begin{aligned} \overline{y}_{11} &= -\overline{y}_{12}\frac{\sigma_2(x_{12} - \alpha_1 x_{32})}{\sigma_1(x_{11} - \alpha_1 x_{31})} \\ \overline{y}_{21} &= -\overline{y}_{22}\frac{\sigma_2(x_{12} - \alpha_1 x_{32})}{\sigma_1(x_{11} - \alpha_1 x_{31})} \\ \overline{y}_{31} &= -\overline{y}_{32}\frac{\sigma_2(x_{12} - \alpha_1 x_{32})}{\sigma_1(x_{11} - \alpha_1 x_{31})} \\ \overline{y}_{41} &= -\overline{y}_{42}\frac{\sigma_2(x_{12} - \alpha_1 x_{32})}{\sigma_1(x_{11} - \alpha_1 x_{31})}. \end{aligned}$$

This set of equations implies a relationship between Y_1 and Y_2 of the form $Y_1 = kY_2$ where $k = \frac{\sigma_2(x_{12}-\alpha_1 x_{32})}{\sigma_1(x_{11}-\alpha_1 x_{12})}$. However a linear dependence between Y_1 and Y_2 is impossible because the columns and rows of Y are orthogonal to one another. This inconsistency is satisfied only for $x_{11} = \alpha_1 x_{31}$ and $x_{12} = \alpha_1 x_{32}$. Relationships between the remaining elements of X_1 and Y_1 may be developed through a similar process. The second and fourth rows of 5.4 have a ratio $\alpha_2 = d_2/d_4$ requiring that $x_{21} = \alpha_2 x_{41}$ and $x_{22} = \alpha_2 x_{42}$. Finally, the ratio of the first and second columns of Equation 5.4 is $\alpha_3 = \frac{d_2 p_{11}}{d_1 p_{12}}$ and the ratio of the third and fourth columns of Equation 5.4 is $\alpha_4 = \frac{d_4 p_{21}}{d_3 p_{22}}$. These relationships between the columns of Equation 5.4 require that $y_{11} = \alpha_3 y_{21}$ and $y_{31} = \alpha_4 y_{41}$. Summarizing these results gives

$$X_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ x_{11}/\alpha_1 \\ x_{21}/\alpha_2 \end{bmatrix}, \quad Y_1 = \begin{bmatrix} y_{11} \\ y_{11}/\alpha_3 \\ y_{31} \\ y_{31}/\alpha_4 \end{bmatrix} \tag{5.5}$$

so that only two independent terms (x_{11}, x_{21} in X_1, and y_{11}, y_{31} in Y_1) appear in each vector. From Section 4.1, equality between the left-hand and right-hand sides of Equation 5.1 occurs only when MPDA conditions are established. This in turn requires pair-wise equality between the elements of X_1 and Y_1 or

$$|x_{i1}| = |y_{i1}| \quad i = 1, \ldots, n^2 \tag{5.6}$$

for similarity scaling with simple $\overline{\sigma}(DM_aD^{-1})$. Applying the requirements of

Equation 5.6 to Equation 5.5 gives

$$\begin{aligned} |x_{11}| &= |y_{11}| \\ |x_{21}| &= \frac{|y_{11}|}{\alpha_3} = \frac{|x_{11}|}{\alpha_3} \\ \frac{|x_{11}|}{\alpha_1} &= |y_{31}| \\ \frac{|x_{21}|}{\alpha_2} &= \frac{|y_{31}|}{\alpha_4} = \frac{|x_{11}|}{\alpha_1\alpha_4} = \frac{|y_{11}|}{\alpha_2\alpha_3}. \end{aligned}$$

From this series of relationships the following equality must hold under MPDA conditions:

$$\frac{\alpha_1\alpha_4}{\alpha_2\alpha_3} = 1. \tag{5.7}$$

Replacing the $\alpha_{i's}$ in Equation 5.7 with the corresponding $d_{i's}$ and $p_{ij's}$ gives

$$\frac{d_2^2 d_3^2 p_{11} p_{22}}{d_1^2 p_{12} p_{21}} = d_4^2$$

which shows that d_4 can be expressed as a simple function of d_1, d_2, d_3 and the $p_{ij's}$. The proof for the 2×2 case is completed by noting that one of the d_i can be always be scaled to "1" without loss of generality so that the total number of free optimization variables is $2 = 2(n-1)$.

Having completed the proof for the 2×2 case, the proof for the general $n \times n$ case follows in the same manner. For $M \in \mathcal{C}^{n\times n}$ the rank of expanded matrix M_a must be less than or equal to n requiring that $\sigma_{n+1}, \ldots, \sigma_{n^2} = 0$. Continuing with the process of equating rows and columns of Equations 5.2 and 5.4 produces relationships between X_1 and Y_1 similar to those of Equation 5.5 when MPDA is established. The number of independent vector elements in X_1 and Y_1 becomes n for the general case and $(n-1)^2$ expressions in the form of Equation 5.7 describe the relationships between

the $2n(n-1)$ $\alpha_{i's}$. From these expressions a general equation relating the dependent and independent values of optimal scaling matrix D can be written as

$$\begin{aligned}
d_i^2 &= \frac{d_{i-n}^2 d_{n+1}^2 p_{11} p_{2,(i-n)}}{d_1^2 p_{(1,(i-n))} p_{2,1}} \qquad (5.8)\\
i &= n+2,\ldots,2n;\\
d_j^2 &= \frac{d_{j-2n}^2 d_{2n+1}^2 p_{11} p_{3,(j-2n)}}{d_1^2 p_{(1,(j-2n))} p_{3,1}}\\
j &= 2n+2,\ldots,3n;\\
\vdots &= \qquad \vdots\\
d_k^2 &= \frac{d_{k-(n-1)n}^2 d_{(n-1)n+1}^2 p_{11} p_{(n,k-(n-1)n)}}{d_1^2 p_{(1,k-(n-1)n)} p_{n,1}}\\
k &= (n-1)n+2,\cdots,n^2.
\end{aligned}$$

Subtracting the $(n-1)^2$ expressions in the form of Equation 5.7 from the n^2-1 variables required for the unreduced scaling structure gives the desired result of

$$n^2 - 1 - (n-1)^2 = 2(n-1)$$

optimization variables. ■

The complexity of the indices in Equation 5.8 unfortunately obscures the simplicity of its application. The following example will hopefully clarify these results.

Example 5.1: Let randomly generated matrices $M \in \mathcal{C}^{3\times3}$ and $P \in \Re^{3\times3}$ be chosen as in Example 4.1 such that

$$M = \begin{bmatrix} .063+.156\jmath & -.322+.480\jmath & .585+.526\jmath \\ .726-.514\jmath & -.323-.344\jmath & .150-.469\jmath \\ .189-.463\jmath & .053-.577\jmath & -.236-.056\jmath \end{bmatrix}$$

and

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} = \begin{bmatrix} 2.99 & 3.03 & 0.54 \\ 1.65 & 1.87 & 3.41 \\ 1.90 & 1.20 & 1.37 \end{bmatrix}.$$

Then expanding $M_a = E_2 M E_1 P_d$ gives $\mu(M_a) = 8.25$ with optimal scaling matrix

$$\hat{D} = \begin{bmatrix} d_1 = 1 \\ d_2 = 0.906 \\ d_3 = 0.406 \\ d_4 = 0.719 \\ d_5 = 0.688 = \sqrt{\frac{d_2^2 d_4^2 p_{11} p_{22}}{d_1^2 p_{12} p_{21}}} \\ d_6 = 0.988 = \sqrt{\frac{d_3^2 d_4^2 p_{11} p_{23}}{d_1^2 p_{13} p_{21}}} \\ d_7 = 0.783 \\ d_8 = 0.560 = \sqrt{\frac{d_2^2 d_7^2 p_{11} p_{32}}{d_1^2 p_{12} p_{31}}} \\ d_9 = 0.636 = \sqrt{\frac{d_3^2 d_7^2 p_{11} p_{33}}{d_1^2 p_{13} p_{31}}} \end{bmatrix}.$$

Of course this is identical to scaling matrix $\hat{D}$ from Example 4.1 found by optimizing over $n^2 - 1$ variables. In this case, the added complexity of computing the dependent elements of D is more than offset by the savings in computations resulting from the reduced number of optimization variables.

Tables 5.1 through 5.3 summarize the savings in floating point operations (flops) due to the scaling method of this section. The "Size" columns represent the dimension of expanded matrix M_a while the number of optimization variables (n^2-1, or $2(n-1)$) column distinguishes between the old and new scaling methods respectively. Both optimization methods were implemented using a MATLAB^{TM} based quasi-Newton method with analytic derivatives [32, 33]. Also, the starting point $D = I$ was chosen for both methods to make the comparison as equitable as possible. This initial choice for D is probably as good as any other for randomly generated M and P matrices. However, stability analysis of an actual system using a frequency sweep could use the optimum $\hat{D}$ from a previous frequency point as an improved initial point.

For the 4×4 systems, the average flop count required to compute μ using $n^2 - 1$

Table 5.1: FLOP Count Comparison for 4×4 Systems

System	Size	Optimization Variables		% Difference
		$n^2 - 1$	$2(n-1)$	
1	4	85,847	70,669	-17.7
2	4	84,463	55,117	-34.7
3	4	102,588	95,816	-6.6
4	4	83,236	74,326	-10.7
5	4	83,239	62,915	-24.4
6	4	67,001	57,630	-14.0
7	4	75,247	61,280	-18.6
8	4	62,237	73,254	+17.7
9	4	78,078	43,474	-43.6
10	4	61,244	58,959	-3.7
11	4	77,018	34,567	-55.1
12	4	51,155	48,680	-4.8
13	4	88,406	73,190	-17.2
14	4	53,459	46,563	-14.7
15	4	57,321	46,417	-19.0
16	4	59,942	49,500	-17.4
17	4	47,572	33,124	-30.3
18	4	56,281	51,606	-8.3
19	4	76,508	61,132	-20.1
20	4	89,680	54,088	-39.7
21	4	63,493	68,793	+8.3
22	4	75,972	58,982	-22.4
23	4	74,279	101,329	+36.4
24	4	77,029	53,882	-30.04
25	4	76,402	39,261	-48.6
Average:		75,637	57,955	-23.4

Table 5.2: FLOP Count Comparison for 9×9 Systems

System	Size	Optimization Variables		% Difference
		$n^2 - 1$	$2(n-1)$	
1	9	1,178,572	1,240,971	5.3
2	9	1,188,389	600,863	-49.4
3	9	1,298,812	991,019	-23.7
4	9	1,079,390	810,888	-24.9
5	9	1,146,836	785,305	-31.5
6	9	1,924,664	1,315,678	-31.6
7	9	2,320,320	1,172,074	-49.5
8	9	1,596,230	1,332,371	-16.5
9	9	1,356,005	1,182,523	-12.8
10	9	1,591,708	1,294,051	-18.7
11	9	772 825	606,518	-21.5
12	9	784,541	922,383	+17.5
13	9	897,372	798,991	-11.0
14	9	1,093,402	1,090,522	-5.8
15	9	906,052	501,610	-44.6
16	9	824,228	847,159	+2.8
17	9	820,406	731,396	-10.8
18	9	1,045,708	865,101	-17.3
19	9	897,115	605,850	-32.5
20	9	1,538,927	982,515	-36.2
21	9	1,366,940	938,130	-31.5
22	9	932,012	722,522	-22.5
23	9	1,190,635	581,242	-51.2
24	9	943,404	990,477	+4.9
25	9	1,270,973	831,299	-17.5
Average:		1,167,769	909,658	-22.0

Table 5.3: FLOP Count Comparison for 16×16 Systems

System	Size	Optimization Variables n^2-1	Optimization Variables $2(n-1)$	% Difference
1	16	8,124,858	5,633,607	-30.7
2	16	9,016,172	4,483,259	-50.3
3	16	10,886,765	9,511,034	-12.6
4	16	8,247,534	6,271,272	-24.0
5	16	8,519,069	5,544,709	-34.9
6	16	9,450,057	5,919,971	-37.4
7	16	10,789,909	9,097,814	-15.7
8	16	9,574,741	8,900,847	-7.0
9	16	10,896,836	8,523,835	-21.8
10	16	11,425,524	7,454,489	-34.8
11	16	8,525 125	6,322,816	-25.8
12	16	8,514,427	5,807,702	-31.8
13	16	9,030,860	8,373,217	-7.3
14	16	7,637,353	5,854,878	-23.3
15	16	10,589,478	5,854,878	-43.9
16	16	10,629,956	6,544,465	-38.4
17	16	9,661,623	8,417,042	-12.9
18	16	9,175,808	5,742,484	-37.4
19	16	7,354,233	6,232,710	-15.3
20	16	7,611,841	5,889,917	-22.6
21	16	10,710,469	6,438,349	-39.9
22	16	11,214,534	9,057,426	-19.2
23	16	7,954,635	6,603,402	-17.0
24	16	9,350,555	6,028,401	-35.5
25	16	8,541,532	7,901,464	-7.5
Average:		8,996,696	6,896,279	-23.3

variables is 75,637 versus 57,955 using $2(n-1)$ variables: a savings of 23.4%. As system size increases, the flop count savings do not change appreciably with an average savings of 22.0% for the 9×9 systems and 23.3% for the 16×16 systems. Since μ must be determined over a wide frequency range for each transfer function, the savings offered by the scaling method of Theorem 5.1 promise substantial reductions in computational effort for the design of robust, multivariable control systems. Note that in some of the examples shown the reduced scaling method actually requires more flops than the full optimization over $n^2 - 1$ variables. This occasional anamoly is probably due to the initial guess of $D = I$. For the unreduced case, all diagonal elements of D may be set to 1, while the reduced structure only allows the free elements to be set to 1. For a few of the examples, having all diagonal elements of $D = 1$ apparently provides the unreduced method with a superior initial guess for the optimization.

Reducing the number of optimization variables offers more than the savings in flop counts. Depending on the minimization algorithm chosen, a possibly substantial savings in memory requirements results as well. For example, the popular quasi-Newton (such as that used in this work) routines require storage on the order of N^2 where N represents the number of optimization variables [34]. Along with the memory requirements comes additional bookkeeping to keep track of the gradient information for the N variables. Therefore the reduction from $n^2 - 1$ to $2(n-1)$ free variables offers computational savings that may not show up directly in flop count comparisons.

5.2 Effects of Zero Elements in P

Example 5.1 and the systems of Table 5.1 represent the maximum reduction of free variables from $n^2 - 1$ to $2(n-1)$. If elements of Δ are replaced with zeros, this efficiency gap begins to diminish although the new method is always at least as efficient in terms of optimization variables as previous similarity scaling methods. In fact, the efficiency advantage holds until the number of zeros equals $n(n-2)+1$ so, for example, a 10×10 system with up to 80 zero elements in Δ will still maintain an advantage in terms of free variables. Of course the relationships shown in Equation 5.8 must be altered to account for the reduction in the d_i's. This is demonstrated in Example 5.2.

Example 5.2: Starting with the same random M and P matrices from Example 4.1, set p_{22} and p_{33} to zero. Then the size of expanded matrix M_a need only increase to 7×7 instead of 9×9 since the zero elements of P can be eliminated. The optimal D for this example is then found to be

$$D = \begin{bmatrix} d_1 = 1 \\ d_2 = 0.852 \\ d_3 = 0.387 \\ d_4 = 0.746 \\ d_5 = 0.976 = \sqrt{\frac{d_3^2 d_4^2 p_{11} p_{23}}{d_1^2 p_{13} p_{21}}} \\ d_6 = 0.868 \\ d_7 = 0.584 = \sqrt{\frac{d_2^2 d_6^2 p_{11} p_{32}}{d_1^2 p_{12} p_{31}}} \end{bmatrix}$$

with $\mu(M_a) = 6.64$. Note that while the expression for d_5 corresponds to that of d_6 in Example 5.1, the expression for d_7 changes slightly to account for the reduction in the number of d_i's. This example shows the reduction from $n^2 - 1 = 8$ to $n^2 - 3 = 6$

for previous methods of similarity scaling. The new method, however, experiences no additional reduction and remains at $2(n-1) = 4$.

5.3 Application to Block Structured Uncertainties

As discussed in Section 2.4, block structured uncertainties represent an important class of uncertainties that, in general, cannot be addressed using nonsimilarity scaling techniques. Therefore, one of the more important advantages of similarity scaling is the ease with which general block structured uncertainties may be handled. Furthermore, by examining the relationships between the elements of X_1 and Y_1 as in Equation 5.5, it is also possible to reduce the number of optimization variables for blocks of arbitrary dimension. Theorem 5.2 follows directly from Theorem 5.1 and the MPDA relationships for block structured uncertainties discussed in Section 4.2.

Theorem 5.2: Given $M \in C^{n\times n}$ and $\Delta \in C^{n\times n}$, where Δ is composed of blocks of dimension less than n, the solution for $\mu(M_b)$ in the optimization problem

$$\mu(M_b) \leq \inf_{D_b} \overline{\sigma}(D_b M_a D_b^{-1}) \tag{5.9}$$

with simple $\overline{\sigma}$ can be found with no more than $2(n-1)$ free variables.

Proof: The proof is straightforward given Theorem 5.1. Recall from Section 3.4 that for block structured uncertainties, scaling matrix D_b contains block-diagonal elements corresponding to the structure of P. For P composed solely of 1×1 blocks, Theorem 5.1 reveals that no more than $2(n-1)$ optimization variables are required to compute $\mu(M_b)$. At the other extreme, for P composed of a single, diagonal, block structured uncertainty of size n, no optimization variables would be required since

the system effectively simplifies to an unstructured uncertainty problem. For systems falling between these limiting cases, singular vector relationships similar to those of Equation 5.5 still apply so the number of optimization variables required to solve any block structured problem cannot exceed $2(n-1)$. ■

Example 5.3 illustrates the application of the optimization variable reduction method to block structured uncertainties.

Example 5.3: Using the same M and Δ matrices from Example 4.2, form expanded matrix M_b followed by DM_bD^{-1}. The rank of DM_aD^{-1} must be less than or equal to 3 so comparing the terms of DM_aD^{-1} in the form of Equation 5.2 with that in the form of Equation 5.4 requires the structure of X_1 and Y_1 to be

$$X_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \\ x_{11}\frac{d_4}{d_1} \\ x_{11}\frac{d_5}{d_1} \\ x_{21}\frac{d_6}{d_2} \\ x_{31}\frac{d_6}{d_3} \end{bmatrix}, \quad Y_1 = \begin{bmatrix} y_{11} \\ y_{11}\frac{p_{12}d_1}{p_{11}d_2} \\ y_{11}\frac{p_{13}d_1}{p_{11}d_3} \\ y_{41} \\ y_{51} \\ y_{41}\frac{p_{22}d_4}{p_{21}d_6} \\ y_{51}\frac{p_{33}d_5}{p_{31}d_6} \end{bmatrix}. \tag{5.10}$$

As in Example 4.2 the optimizing D and the absolute values of the elements of X_1 and Y_1 are

$$\hat{D} = \begin{bmatrix} d_1 = 1 \\ d_2 = 0.82 \\ d_3 = 0.41 \\ d_4 = 0.63 \\ d_5 = 0.87 \\ d_6 = 0.69 \\ d_7 = 0.69 \end{bmatrix}, \quad |X_1| = \begin{bmatrix} 0.427 \\ 0.527 \\ 0.187 \\ 0.269 \\ 0.373 \\ 0.442 \\ 0.313 \end{bmatrix}, \quad |Y_1| = \begin{bmatrix} 0.427 \\ 0.527 \\ 0.187 \\ 0.269 \\ 0.373 \\ 0.279 \\ 0.465 \end{bmatrix}$$

with $\mu(M_b) = 6.5$. Applying the MPDA property for block structures along with the relationships of Equation 5.10 gives a simple expression for d_6 (and d_7) in terms of the other $d_{i's}$ and the elements of P of the form

$$d_6^4 = \frac{\left[\frac{d_4^2 p_{22}}{d_1 p_{21}}\right]^2 + \left[\frac{d_5^2 p_{33}}{d_1 p_{31}}\right]^2}{\left[\frac{d_1 p_{12}}{d_2^2 p_{11}}\right]^2 + \left[\frac{d_1 p_{13}}{d_3^2 p_{11}}\right]^2} = 0.2267. \tag{5.11}$$

Knowing that d_6 is dependent on the elements of P and the other $d_{i's}$ allows the number of optimization variables required for this problem to be reduced from 5 to only 4. Variable reduction for other block structured uncertainties may be determined in an analogous manner and will depend on the number and size of individual blocks. Although Equation 5.11 is more complicated than the relationships for scalar uncertainties, the elimination of unnecessary free variables should produce a net savings in floating point operations: particularly for systems of high order.

5.4 Complete Solution to the Block 2×2 Problem

In the original paper introducing the structured singular value concept, Doyle proved that for $k \leq 3$, where k denotes the number of uncertainty blocks in Δ, the inequality of

$$\mu(M_b) = \sup_{U_b} \rho(M_b U_b) \leq \inf_{D_b} \overline{\sigma}(D_b M_b D_b^{-1}) \tag{5.12}$$

is guaranteed to attain equality at the "inf" regardless of whether or not $\overline{\sigma}$ is simple [15]. For the special 2×2 scalar case where Δ has the form

$$\Delta = \begin{bmatrix} \delta_1 & \delta_2 \\ \delta_3 & \delta_4 \end{bmatrix},$$

Daniel, Kouvaritakis and Latchman have shown that equality of Equation 5.12 may still be guaranteed even though $k = 4$ [31]. The proof for this result depends on formulating the problem in a nonsimilarity scaling context and unfortunately does not extend directly to handle block structured systems. However, by applying the reduced scaling structure of the previous sections to the original proof by Doyle, it is shown that the inequality of Equation 5.12 always becomes equality for any block structured uncertainty of the form

$$\Delta = \begin{bmatrix} \Delta_1 & \Delta_2 \\ \Delta_3 & \Delta_4 \end{bmatrix}.$$

Doyle's original proof guaranteeing equality of Equation 3.22 for $k \leq 3$ requires only that the number of independent optimization variables be less than or equal to two [15]. Since one of the elements of scaling matrix D may always be fixed at some arbitrary value without loss of generality, an uncertainty structure with three or less blocks requires no more than two independent variables for the optimization of $\inf_D \overline{\sigma}(DM_aD^{-1})$. By proving that a 2×2 block structured uncertainty actually requires no more than two optimization variables, this result may be extended to guarantee that $\mu(M_b) = \inf_{D_b} \overline{\sigma}(D_bM_bD_b^{-1})$ for general 2×2 block structured systems.

The following discussion briefly outlines the work of Doyle leading up to his lemma regarding the equality of Equation 5.12 for two or less optimization variables. This outline is presented only as a summary and the reader is referred to [15] for a complete treatment of this material.

Before reviewing Doyle's work, a number of terms must first be defined start-

ing with the notion of a directional derivative for $\overline{\sigma}(D_b M_b D_b^{-1})$ with respect to the elements of D. Denote the multiplicity of $\overline{\sigma}$ as q and decompose $(D_b M_b D_b^{-1})$ as

$$(D_b M_b D_b^{-1}) = \overline{\sigma}(D_b M_b D_b^{-1}) X_a Y_a^H + X_b \Sigma_b Y_b^H$$

where X_a, $Y_a \in \mathcal{C}^{n \times q}$ for $M_b \in \mathcal{C}^{n \times n}$. Next, define some H_j such that

$$H_j = H_j^H = \mathrm{Re}(X_a^H (D_b M_b D_b^{-1}) Y_a).$$

A directional derivative, ∇_2, may then be written as

$$\nabla_2 = \{x \in \Re^r, x_j = v^H H_j v \mid v \in \mathcal{C}^r, v^H v = 1\}$$

where r represents the number of elements in the directional derivative and similarly the number of optimization variables. Formulated in this manner, ∇_2 simply reduces to an ordinary gradient for $q = 1$. Also, for $\nabla_2 = 0$ regardless of the multiplicity of $\overline{\sigma}$, a true stationary point is achieved. From the MPDA theory this stationary point assures that $\mu(M_b) = \inf_D \overline{\sigma}(D_b M_b D_b^{-1})$.

The convex hull of ∇_2, denoted $\mathrm{co}\nabla_2$, also provides gradient-type information and for $q = 1$, $\mathrm{co}\nabla_2$ likewise reduces to an ordinary gradient. However, for $q > 1$, a value of $\mathrm{co}\nabla_2 = 0$ indicates only that the "inf" of $\inf_{D_b} \overline{\sigma}(D_b M_b D_b^{-1})$ has been achieved with no guarantee of a corresponding stationary point. Such a situation may actually be a "cusp" where $\mu(M_b) < \inf_{D_b}(D_b M_b D_b^{-1})$.

Now let $H_j = H_j^H \in \mathcal{C}^{q \times q}$, $j = 1, \ldots, r$ and define some $f : \mathcal{C}^q \rightarrow \Re^r$ as

$$f_j(x) = x^H H_j x \quad \text{for } x \in \mathcal{C}^q.$$

Also, for some positive integer m, let

$$P^m = \{x \in \mathcal{C}^m \mid x^H x = 1\} \subset \mathcal{C}^m$$

and

$$S^m = \{x \in \Re^{m+1} \mid x^T x = 1\} \subset \Re^{m+1}.$$

Using these definitions, the following lemma shows that in certain cases, $\nabla_2 = \text{co}\nabla_2$. In such cases, the condition $\text{co}\nabla_2 = 0$ implies that $\nabla_2 = 0$ and therefore signifies that a stationary point has been found.

Lemma 5.1: For $f : \mathcal{C}^q \rightarrow \Re^r$ defined as above, if $r = 1$ or 2, then $f(P^q)$ is convex.

Proof: See [15].

Lemma 5.1 implies that regardless of the multiplicity of $\overline{\sigma}$, if no more that 2 optimization variables are required to solve Equation 5.12, then $\nabla_2 = \text{co}\nabla_2$ and the minimum of the infimization must be a stationary point and hence equals $\mu(M_b)$. The following theorem reveals that no more than 2 free optimization variables are required to solve Equation 5.12 for systems with 2×2 block structured uncertainties allowing the results of Lemma 5.1 to be invoked.

Theorem 5.3: For the case of 2×2 block structured uncertainties, no more than two optimization variables are required to solve $\inf_{D_b} \overline{\sigma}(D_b M_b D_b^{-1})$ regardless of the individual block sizes.

Proof: Define $M \in \mathcal{C}^{n \times n}$ and 2×2 uncertainty matrix Δ with corresponding P matrix as

$$\Delta = \begin{bmatrix} \Delta_1 & \Delta_2 \\ \Delta_3 & \Delta_4 \end{bmatrix} \rightarrow P = \begin{bmatrix} P_1 \cdot I_m & P_2 \cdot I_m \\ P_3 \cdot I_m & P_4 \cdot I_m \end{bmatrix}$$

with $P \in \Re^{n \times n}$, $P_i \in \Re^{m \times m}$ and $m = n/2$. Then similarity scaling matrix D has the

structure

$$D = diag[\underbrace{d_1,\ldots,d_1}_{m},\ \underbrace{d_2,\ldots,d_2}_{m},\ \underbrace{d_3,\ldots,d_3}_{m},\ \underbrace{d_4,\ldots,d_4}_{m}].$$

Expanding the system to the form of Equations 5.2 and 5.4 and comparing terms requires the elements of X_1 and Y_1 to be

$$X_1 = \begin{bmatrix} x_{1,1} \\ \vdots \\ x_{m,1} \\ x_{(m+1),1} \\ \vdots \\ x_{2m,1} \\ x_{1,1}\frac{d_3}{d_1} \\ \vdots \\ x_{m,1}\frac{d_3}{d_1} \\ x_{(m+1),1}\frac{d_4}{d_2} \\ \vdots \\ x_{2m,1}\frac{d_4}{d_2} \end{bmatrix}, \quad Y_1 = \begin{bmatrix} y_{1,1} \\ \vdots \\ y_{m,1} \\ y_{1,1}\frac{\overline{\sigma}(P_2)d_1}{\overline{\sigma}(P_1)d_2} \\ \vdots \\ y_{m,1}\frac{\overline{\sigma}(P_2)d_1}{\overline{\sigma}(P_1)d_2} \\ y_{(2m+1),1} \\ \vdots \\ y_{(3m),1} \\ y_{(2m+1),1}\frac{\overline{\sigma}(P_4)d_3}{\overline{\sigma}(P_3)d_4} \\ \vdots \\ y_{(3m),1}\frac{\overline{\sigma}(P_4)d_3}{\overline{\sigma}(P_3)d_4} \end{bmatrix}. \tag{5.13}$$

Applying the MPDA requirement for block structured uncertainties to Equation 5.13 gives a relationship between the elements of D as

$$d_4 = \frac{d_2 d_3}{d_1}\sqrt{\frac{\overline{\sigma}(P_1)\overline{\sigma}(P_4)}{\overline{\sigma}(P_2)\overline{\sigma}(P_3)}}.$$

Fixing $d_1 = 1$ leaves only d_2 and d_3 as independent variables completing the proof. ■

Example 5.4: As a simple example, consider the following 4×4 problem with four 2×2 block structured uncertainties.

$$M = \begin{bmatrix} 0.8+j0.3 & 1.0+j0.9 & 0.1+j0.5 & 0.4+j0.5 \\ 1.0+j0.4 & 0.7+j0.9 & 0.6+j0.5 & 0.8+j0.3 \\ 0.4+j0.2 & 0.8+j0.1 & 0.9+j0.3 & 0.5+j0.1 \\ 0.2+j0.5 & 0.7+j0.9 & 0.3+j1.0 & 0.2+j0.9 \end{bmatrix}$$

with

$$\Delta = \begin{bmatrix} \Delta_1 & \Delta_2 \\ \Delta_3 & \Delta_4 \end{bmatrix} \rightarrow P = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 3 & 0 & 4 & 0 \\ 0 & 3 & 0 & 4 \end{bmatrix}.$$

The singular vector relationships of Equation 5.13 become

$$X_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \\ x_{41} \\ x_{11}\frac{d_3}{d_1} \\ x_{21}\frac{d_3}{d_1} \\ x_{31}\frac{d_4}{d_2} \\ x_{41}\frac{d_4}{d_2} \end{bmatrix}, \quad Y_1 = \begin{bmatrix} y_{11} \\ y_{21} \\ y_{11}\frac{\overline{\sigma}(P_1)d_2}{\overline{\sigma}(P_2)d_1} \\ y_{21}\frac{\overline{\sigma}(P_1)d_2}{\overline{\sigma}(P_2)d_1} \\ y_{51} \\ y_{61} \\ y_{51}\frac{\overline{\sigma}(P_4)d_3}{\overline{\sigma}(P_3)d_4} \\ y_{61}\frac{\overline{\sigma}(P_4)d_3}{\overline{\sigma}(P_3)d_4} \end{bmatrix}.$$

For this particular example, the optimal D is

$$D = \begin{bmatrix} d_1 = 1 \\ d_1 = 1 \\ d_2 = 1.38 \\ d_2 = 1.38 \\ d_3 = 1.65 \\ d_3 = 1.65 \\ d_4 = 1.86 = \frac{d_2 d_3}{d_1}\sqrt{\frac{(1)(4)}{(2)(3)}} \\ d_4 = 1.86 = \frac{d_2 d_3}{d_1}\sqrt{\frac{(1)(4)}{(2)(3)}} \end{bmatrix}$$

with $\mu(M_b) = 16.43$.

5.5 Convergence Properties of the Reduced Scaling Structure

The previous sections show that when μ is found, the corresponding optimal scaling matrix D exhibits a degree of redundancy which allows a reduction in the number of independent optimization variables. It is, however, important to ensure that when the relationships between the $d_{i'}$s and $p_{i'j'}$s are incorporated in a reduced scaling structure from the outset, the convexity properties of the original problem

are not lost. The remainder of this section will show that, in fact, these important convexity properties do still apply. Furthermore, MPDA conditions used to establish equality of Equation 3.22 for the full scaling structure also pertain to the reduced scaling structure guaranteeing its convergence to μ.

Convexity of $\overline{\sigma}^2(DM_aD^{-1})$ in terms of e^D, $D \in \mathcal{D}$ was first proven by Safonov and Doyle [28]. Subsequent authors have offered alternative proofs in somewhat simpler forms [35, 29]. The following lemma concerning the convexity of some function f was proven by Tsing [36].

Lemma 5.2: Suppose $\forall x$, g_x: $\Re \mapsto \Re$ twice differentiable such that $f(x) = g_x(x)$, $f(t) \geq g_x(t)$ and $\frac{d^2}{dt^2} g_x(t)|_{t=x} \geq 0$. Then f is convex.

Proof: See [36].

Lemma 5.2 may then be applied to the scaled singular value problem to prove convexity of $\overline{\sigma}(e^D M_a e^{-D})$. For notational brevity, denote $M_x = e^{Dx} M_a e^{-Dx}$ with singular vectors u and v such that $f(x) = \overline{\sigma}(M_x) = u^H M_x v$. Next, define $g_x(t) = \Re\{u^H e^{Dt} M_a e^{-Dt} v\}$. Since $f(t) \geq g_x(t)$ and

$$\begin{aligned}
\frac{d^2}{dt^2} g_x(t)|_{t=x} &= \Re\{u^H (D^2 M_x - 2DM_xD + M_xD^2)v\} \\
&= f(x)(u^H D^2 u + v^H D^2 v) - 2\Re\{u^H D M_x D v\} \\
&= [u^H D \;\; v^H D] \begin{bmatrix} f(x)I & -M_x \\ -M_x^H & f(x)I \end{bmatrix} \begin{bmatrix} Du \\ Dv \end{bmatrix} \geq 0.
\end{aligned}$$

Therefore, by Lemma 5.2 f is convex.

This convexity property guarantees that any local minimum of $\overline{\sigma}(DM_aD^{-1})$ with respect to D is always a global minimum. For a new scaling matrix, S, that inherently

incorporates the reduced scaling structure of Equation 5.8, the convexity lemma still applies since S is a subset of $\mathcal{D}$. Therefore, it must be shown that the value of $\overline{\sigma}(SM_aS^{-1})$ at the minimum always corresponds to μ and not some other value.

Theorem 5.4: For diagonal scaling matrices $D, S \in \Re^{n^2 \times n^2}$ where S conforms to the scaling structure of Equation 5.8, the following equality must hold

$$\inf_D \overline{\sigma}(DM_aD^{-1}) = \inf_S \overline{\sigma}(SM_aS^{-1})$$

whenever a stationary point occurs at the "inf" of either side of the equality.

Proof: This proof requires examination of the stationary points at which $\frac{\partial \overline{\sigma}(SM_aS^{-1})}{\partial S_i} = 0$. Any maxima of $\overline{\sigma}(SM_aS^{-1})$ with respect to S occurs at infinity because $\overline{\sigma}$ increases without bound with the free elements of S. This eliminates the possibility of any stationary point corresponding to a maxima. Inflection points are not possible due to the convexity properties of e^S so any stationary point must coincide with a minima. Again, the convexity property requires any local minimum to be a global minimum so if a stationary point can be found, it must correspond to the unique minimum.

Because the maximum singular value is always a positive function of some complex matrix $A \in \mathcal{C}^{n \times m}$, the stationary points of $\overline{\sigma}^2(A)$ correspond to those of $\overline{\sigma}(A)$. This allows an alternative form of the singular values to be written in terms of the eigenvalues giving $\overline{\sigma}^2(A) = \lambda_1(A^HA)$ where λ_1 represents the eigenvalue of maximum magnitude. Denoting the eigenvector corresponding to λ_1 as W_1, the eigenvalue/eigenvector relationship may be written as

$$A^HAW_1 = \lambda_1 W_1 \tag{5.14}$$

and the derivative with respect to some variable x as

$$\frac{\partial A^H A}{\partial x} W_1 + A^H A \frac{\partial W_1}{\partial x} = \lambda_1 \frac{\partial W_1}{\partial x} + W_1 \frac{\partial \lambda}{\partial x}.$$

Multiplying on the left by W_1^H gives

$$W_1^H \frac{\partial A^H A}{\partial x} W_1 + W_1^H A^H A \frac{\partial W_1}{\partial x} = W_1^H \lambda_1 \frac{\partial W_1}{\partial x} + \frac{\partial \lambda_1}{\partial x}.$$

which can then be simplified to

$$\frac{\partial \lambda_1}{\partial x} = W_1^H \frac{\partial A^H A}{\partial x} W_1 \tag{5.15}$$

giving an expression for the derivative of λ_1 with respect to x. The derivative with respect to some element of scaling matrix S can then be found by replacing the variables of Equation 5.14 such that

$$\begin{aligned} \mathcal{A} &= A^H A = S^{-1} M_a^H D^2 M_a S^{-1} \\ W_1 &= Y_1 \\ \lambda_1 &= \overline{\sigma}^2. \end{aligned}$$

The differentiation will first be carried out for a 2×2 case with respect to element s_{22} of S where

$$S = diag[s_{11},\ s_{22},\ s_{33},\ s_{22}s_{33}\sqrt{\frac{p_{11}p_{22}}{p_{12}p_{21}}}].$$

First decompose S such that

$$S = S_0 + s_{22}E_{22} + s_{22}s_{33}E_{44}k$$

where $S_0 = diag[s_{11},\ 0,\ s_{33},\ 0]$, $k = \sqrt{\frac{p_{11}p_{44}}{p_{22}p_{33}}}$, and the E_{ii} are $n \times n$ matrices with a 1 at the iith position and 0 elsewhere. Likewise, S^{-1} may be decomposed as

$$S^{-1} = S_0^{-1} + \frac{E_{22}}{s_{22}} + \frac{E_{44}}{s_{22}s_{33}k}.$$

Differentiation of $\mathcal{A}$ with respect to s_{22} then proceeds directly with

$$\begin{aligned}\frac{\partial \mathcal{A}}{\partial s_{22}} &= -\frac{E_{22}}{s_{22}^2} M_a^H S^2 M_a S^{-1} - \frac{E_{44}}{s_{22}^2 s_{33} k} M_a^H S^2 M_a S^{-1} \\ &\quad -S^{-1} M_a^H S^2 M_a \left[\frac{E_{22}}{s_{22}^2} + \frac{E_{44}}{s_{22}^2 s_{33} k}\right] \\ &\quad +2 s_{22} S^{-1} M_a^H \left[E_{22} M_a S^{-1} + E_{44} s_{33}^2 k^2 M_a S^{-1}\right].\end{aligned}$$

The E_{ii} terms contain only one nonzero value so this expression may be rewritten as

$$\begin{aligned}\frac{\partial \mathcal{A}}{\partial s_{22}} &= -\frac{s_{22}}{s_{22}^2} E_{22} S^{-1} M_a^H S^2 M_a S^{-1} - \frac{s_{22} s_{33}}{s_{22}^2 s_{33} k} E_{44} S^{-1} M_a^H S^2 M_a S^{-1} \\ &\quad -\frac{s_{22}}{s_{22}^2} S^{-1} M_a^H S^2 M_a S^{-1} E_{22} - S^{-1} M_a^H S^2 M_a S^{-1} E_{44} \frac{s_{22} s_{33} k}{s_{22}^2 s_{33} k} \\ &\quad + 2 s_{22} S^{-1} M_a^H E_{22} M_a S^{-1} + 2 s_{22} s_{33}^2 k^2 S^{-1} M_a^H E_{44} M_a S^{-1}.\end{aligned}$$

Multiplying $\frac{\partial \mathcal{A}}{\partial s_{22}}$ on the left by Y_1^H and the right by Y_1 as in Equation 5.15, gives an expression for $\frac{\partial \overline{\sigma}^2}{\partial s_{22}}$ of

$$\begin{aligned}\frac{\partial \overline{\sigma}^2}{\partial s_{22}} = Y_1^H \frac{\partial \mathcal{A}}{\partial s_{22}} Y_1 &= -2\left[\frac{Y_1^H}{s_{22}} \overline{\sigma}^2 E_{22} Y_1 + \frac{Y_1^H}{s_{22}} \overline{\sigma}^2 E_{44} Y_1\right] \\ &\quad +2 s_{22} Y_1^H \left[S^{-1} M_a^H E_{22} M_a S^{-1} + s_{33}^2 k^2 S^{-1} M_a^H E_{44} M_a S^{-1}\right] Y_1.\end{aligned}$$

Noting that

$$\frac{\partial \overline{\sigma}^2}{\partial s_{22}} = 2\overline{\sigma} \frac{\partial \overline{\sigma}}{\partial s_{22}},$$

and

$$Y_1^H S^{-1} M_a^H S E_{ii} = \overline{\sigma} X_1^H E_{ii},$$

a final expression for $\frac{\partial \overline{\sigma}}{\partial s_{22}}$ may be written as

$$\frac{\partial \overline{\sigma}}{\partial s_{22}} = \frac{\overline{\sigma}}{s_{22}} \left[X_1^H (E_{22} + E_{44}) X_1 - Y_1^H (E_{22} + E_{44}) Y_1\right]. \tag{5.16}$$

Carrying out this process with respect to s_{33} gives an expression similar to Equation 5.16 of

$$\frac{\partial\overline{\sigma}}{\partial s_{33}} = \frac{\overline{\sigma}}{s_{33}}\left[X_1^H(E_{33}+E_{44})X_1 - Y_1^H(E_{33}+E_{44})Y_1\right]. \tag{5.17}$$

Setting Equations 5.16 and 5.17 to zero establishes conditions at which the new scaling technique reaches its global minimum. The stationary point therefore corresponds to the value of S where

$$[\overline{x}_{21}x_{21} - \overline{y}_{21}y_{21} + \overline{x}_{41}x_{41} - \overline{y}_{41}y_{41}] = 0 \tag{5.18}$$

and

$$[\overline{x}_{31}x_{31} - \overline{y}_{31}y_{31} + \overline{x}_{41}x_{41} - \overline{y}_{41}y_{41}] = 0 \tag{5.19}$$

simultaneously.

The MPDA property guarantees that when element-by-element alignment of the principal singular vectors of $\overline{\sigma}(SM_aS^{-1})$ occurs, some U_d exists such that $\overline{\sigma}(SM_aS^{-1}) = \rho(M_aU_d) = \mu(M_a)$. From Equations 5.18 and 5.19, such an alignment of X_1 and Y_1 always coincides with a stationary point of $\frac{\partial\overline{\sigma}(SM_aS^{-1})}{\partial S_i}$ because it requires that

$$|x_{21}|^2 - |y_{21}|^2 = |x_{31}|^2 - |y_{31}|^2 = |x_{41}|^2 - |y_{41}|^2 = 0.$$

Therefore one, and hence the only, minima of $\overline{\sigma}(SM_aS^{-1})$ with respect to S must correspond to $\mu(M_a)$ as long as some stationary point does exist.

Extension to higher order systems is simply a matter of determining which elements of S depend on a particular s_i. For each occurrence, an additional

$$X_1^H E_{jj} X_1 - Y_1^H E_{jj} Y_1$$

expression is included where the jj terms denote the elements of S dependent on a particular s_i term. Again, general $n \times n$ systems similarly invoke MPDA conditions guaranteeing that the stationary point corresponds to $\mu(M_a)$. ■

CHAPTER 6
DIRECT RELATIONSHIP BETWEEN R, L, AND D

6.1 Problem Formulation

The results of Chapter 5 show the number of optimization variables required for similarity scaling to be no more than that required for nonsimilarity scaling. It would therefore seem consistent that some direct relationship exists between scaling matrices D for $\inf_{D\in\mathcal{D}}\overline{\sigma}(DM_aD^{-1})$ and R and L for $\inf_{R,L}\overline{\sigma}(R^{-1}ML^{-1})\overline{\sigma}(LPR)$ since both methods accurately compute μ. The following Theorem presents such a relationship.

Theorem 6.1: For element-by-element structured uncertainties, given some scaling matrix D conforming to the structure of Equation 5.8, a corresponding R and L of the form

$$r_i^2 = \frac{d_1^2 + d_{n+1}^2 + d_{2n+1}^2 \cdots d_{n(n-1)+1}^2}{d_i^2 + d_{n+i}^2 + d_{2n+i}^2 \cdots d_{n(n-1)+i}^2} \quad i = 2,\ldots,n \tag{6.1}$$

and

$$l_i^2 = \frac{\frac{p_{1,1}^2}{d_1^2} + \frac{p_{1,2}^2}{d_2^2} \cdots \frac{p_{1,n}^2}{d_n^2}}{\frac{p_{i,1}^2}{d_{n(i-1)+1}^2} + \frac{p_{i,2}^2}{d_{n(i-1)+2}^2} \cdots \frac{p_{i,n}^2}{d_{n(i-1)+n}^2}} \quad i = 2,\ldots,n \tag{6.2}$$

may be found without further optimization such that

$$\sigma_i(R^{-1}ML^{-1})\overline{\sigma}(LPR) = \sigma_i(DM_aD^{-1}) \quad i = 1,\ldots,n. \tag{6.3}$$

Proof: Rather than dealing explicitly with singular values, it is more convenient in this case to use the equivalent representation

$$\sigma_i(A) = \sqrt{\lambda_i(A^HA)} \quad A \in C^{n\times n}, \quad i = 1,\ldots,n$$

By similarly expanding β, B and C are found to be

$$\begin{aligned} B &= -\overline{M}_{11}M_{11}\left[d_1^2+d_3^2\right]\cdot\left[\frac{p_{11}^2}{d_1^2}+\frac{p_{12}^2}{d_2^2}\right]-\overline{M}_{21}M_{21}\left[d_2^2+d_4^2\right]\cdot\left[\frac{p_{11}^2}{d_1^2}+\frac{p_{12}^2}{d_2^2}\right] \\ &\quad -\overline{M}_{12}M_{12}\left[d_1^2+d_3^2\right]\cdot\left[\frac{p_{21}^2}{d_3^2}+\frac{p_{22}^2}{d_4^2}\right]-\overline{M}_{22}M_{22}\left[d_2^2+d_4^2\right]\cdot\left[\frac{p_{21}^2}{d_3^2}+\frac{p_{22}^2}{d_4^2}\right] \end{aligned}$$

$$\begin{aligned} C &= K\cdot\left[\overline{M}_{11}M_{11}\overline{M}_{22}M_{22}+\overline{M}_{21}M_{21}\overline{M}_{12}M_{12}\right] \\ &\quad -K\cdot\left[\overline{M}_{11}M_{12}\overline{M}_{22}M_{21}+\overline{M}_{11}M_{12}\overline{M}_{21}M_{22}\right] \\ K &= \left[d_1^2+d_3^2\right]\cdot\left[d_2^2+d_4^2\right]\cdot\left[\frac{p_{11}^2p_{21}^2}{d_1^2d_3^2}+\frac{p_{11}^2p_{22}^2}{d_1^2d_4^2}+\frac{p_{12}^2p_{21}^2}{d_2^2d_3^2}+\frac{p_{12}^2p_{22}^2}{d_2^2d_4^2}\right]. \end{aligned}$$

Equating $b\overline{\sigma}^2(LPR) = B$ and $c\overline{\sigma}^4(LPR) = C$, the following direct relationships between R, L, and D will hold

$$r_2^2=\left[\frac{d_1^2+d_3^2}{d_2^2+d_4^2}\right],\qquad l_2^2=\left[\frac{\frac{p_{11}^2}{d_1^2}+\frac{p_{12}^2}{d_2^2}}{\frac{p_{21}^2}{d_3^2}+\frac{p_{22}^2}{d_4^2}}\right] \tag{6.4}$$

but only when equality constraint

$$\overline{\sigma}^2(LPR)=\left[d_1^2+d_3^2\right]\cdot\left[\frac{p_{11}^2}{d_1^2}+\frac{p_{12}^2}{d_2^2}\right] \tag{6.5}$$

applies. At first glance it may appear that the constraint of Equation 6.5 severely limits conditions under which Equation 6.4 holds. The remainder of the proof will reveal, however, that this constraint simply requires the reduced similarity scaling structure developed in Chapter 5.

The left hand side of Equation 6.5 may always be represented as $\lambda_1[(LPR)^H(LPR)]$. Expanding matrix $(LPR)^H(LPR)$ gives

$$(LPR)^H(LPR)=\begin{bmatrix} p_{11}^2+p_{21}^2l^2 & p_{11}p_{12}r_2+p_{21}p_{22}l_2^2r_2 \\ p_{11}p_{12}r_2+p_{21}p_{22}l_2^2r_2 & p_{12}^2r_2^2+p_{22}^2r_2^2l_2^2 \end{bmatrix}.$$

The corresponding characteristic equation of this matrix becomes

$$\lambda^2 + \hat{b}\lambda + \hat{c}.$$

Denote the two roots of Equation 6.7 as $\overline{\sigma}^2(LPR)$ and $\sigma_2(LPR)$. Coefficients $\hat{b}$ and $\hat{c}$ may be expressed in terms of these roots such that

$$\hat{b} = \overline{\sigma}^2(LPR) + \sigma_2^2(LPR) \tag{6.6}$$

$$\hat{c} = \overline{\sigma}^2(LPR) \cdot \sigma_2^2(LPR). \tag{6.7}$$

Relating these expressions to the elements of (LPR) gives

$$\hat{b} = \overline{\sigma}^2(LPR) + \sigma_2^2(LPR) = p_{11}^2 + p_{21}^2 l_2^2 + p_{12}^2 r_2^2 + p_{22}^2 r_2^2 l_2^2 \tag{6.8}$$

and

$$\hat{c} = \overline{\sigma}^2(LPR) \cdot \sigma_2^2(LPR) \tag{6.9}$$

$$= (p_{11}^2 + p_{21}^2 l_2^2) \cdot (p_{12}^2 r_2^2 + p_{22}^2 r_2^2 l_2^2) \tag{6.10}$$

$$- (p_{11}p_{12}r_2 + p_{21}p_{22}l_2^2 r_2) \cdot (p_{11}p_{12}r_2 + p_{21}p_{22}l_2^2 r_2). \tag{6.11}$$

Replacing r_2 and l_2 in Equations 6.8 and 6.11 with the relationships of Equation 6.4 and simplifying gives

$$\hat{b} = p_{11}^2 + p_{12}^2 \left[\frac{d_1^2 + d_3^2}{d_2^2 + d_4^2}\right] + p_{21}^2 \left[\frac{\frac{p_{11}^2}{d_1^2} + \frac{p_{12}^2}{d_2^2}}{\frac{p_{21}^2}{d_3^2} + \frac{p_{22}^2}{d_4^2}}\right] + p_{22}^2 \left[\frac{d_1^2 + d_3^2}{d_2^2 + d_4^2}\right] \cdot \left[\frac{\frac{p_{11}^2}{d_1^2} + \frac{p_{12}^2}{d_2^2}}{\frac{p_{21}^2}{d_3^2} + \frac{p_{22}^2}{d_4^2}}\right]$$

and

$$\hat{c} = \left[d_1^2 + d_3^2\right] \cdot \left[\frac{p_{11}^2}{d_1^2} + \frac{p_{12}^2}{d_2^2}\right] \cdot \frac{\left[p_{11}p_{22} - p_{12}p_{21}\right]^2}{\left[d_2^2 + d_4^2\right] \cdot \left[\frac{p_{21}^2}{d_3^2} + \frac{p_{22}^2}{d_4^2}\right]}. \tag{6.12}$$

Equation 6.7 shows coefficient $\hat{c}$ to be the product of $\overline{\sigma}^2(LPR)$ and $\sigma_2^2(LPR)$. Combining Equations 6.5 and 6.12 then gives the expression for $\sigma_2^2(LPR)$ of

$$\sigma_2^2 = \frac{[p_{11}p_{22} - p_{12}p_{21}]^2}{[d_2^2 + d_4^2]\left[\frac{p_{21}^2}{d_3^2} + \frac{p_{22}^2}{d_4^2}\right]}. \tag{6.13}$$

Equation 6.7 also shows that coefficient $\hat{b}$ is the sum of $\overline{\sigma}^2(LPR)$ and $\sigma_2^2(LPR)$ so that $\hat{b}$ may likewise be written as

$$\hat{b} = \left[d_1^2 + d_3^2\right] \cdot \left[\frac{p_{11}^2}{d_1^2} + \frac{p_{12}^2}{d_2^2}\right] + \frac{[p_{11}p_{22} - p_{12}p_{21}]^2}{[d_2^2 + d_4^2] \cdot \left[\frac{p_{21}^2}{d_3^2} + \frac{p_{22}^2}{d_4^2}\right]}.$$

This gives a new expression for $\hat{b}$ that must be satisfied along with Equation 6.8 so that

$$\begin{aligned}
\hat{b} &= p_{11}^2 + p_{12}^2\left[\frac{d_1^2 + d_3^2}{d_2^2 + d_4^2}\right] + p_{21}^2\left[\frac{\frac{p_{11}^2}{d_1^2} + \frac{p_{12}^2}{d_2^2}}{\frac{p_{21}^2}{d_3^2} + \frac{p_{22}^2}{d_4^2}}\right] + p_{22}^2\left[\frac{d_1^2 + d_3^2}{d_2^2 + d_4^2}\right] \cdot \left[\frac{\frac{p_{11}^2}{d_1^2} + \frac{p_{12}^2}{d_2^2}}{\frac{p_{21}^2}{d_3^2} + \frac{p_{22}^2}{d_4^2}}\right] \\
&= \left[d_1^2 + d_3^2\right] \cdot \left[\frac{p_{11}^2}{d_1^2} + \frac{p_{12}^2}{d_2^2}\right] + \frac{[p_{11}p_{22} - p_{12}p_{21}]^2}{[d_2^2 + d_4^2] \cdot \left[\frac{p_{21}^2}{d_3^2} + \frac{p_{22}^2}{d_4^2}\right]}.
\end{aligned}$$

Multiplying out these two expressions and eliminating common terms reveals that they are equal whenever

$$d_1^4 d_4^4 p_{12}^2 p_{21}^2 + d_2^4 d_3^4 p_{11}^2 p_{22}^2 - 2d_1^2 d_2^2 d_3^2 d_4^2 p_{11} p_{12} p_{21} p_{22} = 0. \tag{6.14}$$

Equation 6.14 can be rewritten as

$$\left[d_1^2 d_4^2 p_{12} p_{21} - d_2^2 d_3^2 p_{11} p_{22}\right]^2 = 0$$

so the constraint of Equation 6.5 will hold for D and P such that

$$d_4 = d_2 d_3 \sqrt{\frac{p_{11}p_{22}}{p_{12}p_{21}}}.$$

This, however, is the same dependent relationship between the $p_{i,j}$'s and d_{i}'s of Equation 5.8 so the direct relationship between R and L of nonsimilarity scaling and D of similarity scaling will always exist whenever the new scaling structure is invoked.

For the $n \times n$ case, a general expression for R and L as a function of D may be derived in an analogous manner by comparing the characteristic equations of the corresponding matrix systems. For $M \in C^{3\times3}$, the characteristic equations would contain three coefficients which must then be matched up as before. While intuitively simple, th[illegible]come too unwieldy to present here. However, [illegible]ation 6.4 may be written as

$$r_i^2 = \frac{\cdot_1 + d_{n+1}^2 + d_{2n+1}^2 \cdots d_{n(n-1)+1}^2}{d_i^2 + d_{n+i}^2 + d_{2n+i}^2 \cdots d_{n(n-1)+i}^2} \quad i = 1 \ldots n$$

and

$$l_i^2 = \frac{\frac{p_{i,1}^2}{d_1^2} + \frac{p_{i,2}^2}{d_2^2} \cdots \frac{p_{i,n}^2}{d_n^2}}{\frac{p_{i,1}^2}{d_{n(i-1)+1}^2} + \frac{p_{i,2}^2}{d_{n(i-1)+2}^2} \cdots \frac{p_{i,n}^2}{d_{n(i-1)+n}^2}} \quad i = 1 \ldots n$$

completing the proof. ■

The following example illustrates the application of these expressions.

Example 6.1: Using M and $\hat{D}$ from Example 4.1, optimal values for R and L are found to be

$$r_2 = \sqrt{\frac{d_1^2 + d_4^2 + d_7^2}{d_2^2 + d_5^2 + d_8^2}} = 1.151, \quad r_3 = \sqrt{\frac{d_1^2 + d_4^2 + d_7^2}{d_3^2 + d_6^2 + d_9^2}} = 1.174$$

and

$$l_2 = \sqrt{\frac{\frac{p_{11}^2}{d_1^2} + \frac{p_{12}^2}{d_2^2} + \frac{p_{13}^2}{d_3^2}}{\frac{p_{21}^2}{d_4^2} + \frac{p_{22}^2}{d_5^2} + \frac{p_{23}^2}{d_6^2}}} = 0.9443, \quad l_3 = \sqrt{\frac{\frac{p_{11}^2}{d_1^2} + \frac{p_{12}^2}{d_2^2} + \frac{p_{13}^2}{d_3^2}}{\frac{p_{31}^2}{d_7^2} + \frac{p_{32}^2}{d_8^2} + \frac{p_{33}^2}{d_9^2}}} = 1.203$$

with

$$\begin{aligned}
\sigma_1(R^{-1}ML^{-1})\overline{\sigma}(LPR) &= \sigma_1(DM_aD^{-1}) = \mu(M_a) = 8.25 \\
\sigma_2(R^{-1}ML^{-1})\overline{\sigma}(LPR) &= \sigma_2(DM_aD^{-1}) = 5.72 \\
\sigma_3(R^{-1}ML^{-1})\overline{\sigma}(LPR) &= \sigma_3(DM_aD^{-1}) = 1.76.
\end{aligned}$$

It should be noted that this transformation not only provides $\hat{L}$ and $\hat{R}$ given $\hat{D}$, but also allows the computation of $\hat{D}$ given $\hat{R}$ and $\hat{L}$. However, this reverse transformation results in a transcendental function that must be solved iteratively. The iterative solution is still quite fast since it involves only elementary math operations: no singular value or eigenvalue decompositions.

6.2 Independence from M

An interesting observation concerning Equations 6.1 and 6.2 is that they are not functions of M. This independence from M requires that any P, R, L, D combination satisfying Equation 6.3 for a given M, must also satisfy Equation 6.3 for any other M. As a simple example consider the following system with two randomly generated M matrices.

$$M_1 = \begin{bmatrix} 0.87 - i2.01 & 4.70 - i0.08 \\ 4.16 - i0.59 & -0.92 + i0.13 \end{bmatrix} \quad M_2 = \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix}$$

$$P = \begin{bmatrix} 1.59 & 1.64 \\ 2.81 & 0.83 \end{bmatrix} \quad D = diag[1,\ 1.415,\ 1.401,\ 1.057]$$

where D satisfies the new scaling structure. Values for R and L are found to be

$$R = diag[1,\ 0.974] \quad L = diag[1,\ 0.912]$$

so that Equation 6.3 becomes

$$\sigma_1(R^{-1}M_1L^{-1})\overline{\sigma}(LPR) = \sigma_1(DM_{a1}D^{-1}) = 20.28$$

$$\sigma_2(R^{-1}M_1L^{-1})\overline{\sigma}(LPR) = \sigma_2(DM_{a1}D^{-1}) = 13.04.$$

Using the same R, L, and D values for M_2 with no optimization gives

$$\sigma_1(R^{-1}M_2L^{-1})\overline{\sigma}(LPR) = \sigma_1(DM_{a2}D^{-1}) = 19.18$$

$$\sigma_2(R^{-1}M_1L^{-1})\overline{\sigma}(LPR) = \sigma_2(DM_{a2}D^{-1}) = 1.34.$$

The direct relationship between R, L and D leads to the following corollary concerning systems that "cusp" during the optimization process.

Corollary 6.1: Systems with element-by-element structured uncertainties with non-simple $\overline{\sigma}(R^{-1}ML^{-1})$ ("cusps") for nonsimilarity scaling must also produce cusps for similarity scaling techniques and vice-versa.

Proof: The proof follows directly from the one-to-one correspondence between the $\sigma_{i's}$ of $\sigma_i(R^{-1}ML^{-1})\overline{\sigma}(LPR)$ and $\sigma_i(DM_aD^{-1})$ required by Equation 6.3. ■

Corollary 6.1 serves to eliminate the appealing notion that a one of these two scaling techniques may converge to μ even if the other method fails. While some alternative scaling technique may eventually overcome the difficulties associated with cusping systems it is the case for nonsimilarity or similarity scaling.

Besides ruling out the use of an alternative scaling method for cusping systems, the direct relationship between R, L, and D serves to improve the vector optimization method of Fan and Tits. The next chapter includes a discussion of this method and the improvements offered by the R, L, D relationship.

CHAPTER 7
REDUCTION OF OPTIMIZATION VARIABLES REQUIRED FOR THE METHOD OF FAN AND TITS

7.1 Constrained Vector Optimization

While the scaling structure developed in Chapter 5 allows a reduction in the number of optimization variables required to compute μ, the process still involves a large number of singular value decompositions. To avoid these computationally intensive decompositions, Fan and Tits introduced a constrained vector optimization procedure that calculates μ using a series of vector norms [37]. The problem formulation is similar to the eigenvalue/eigenvector relationship for matrix $A \in C^{n\times n}$

$$AW_i = \lambda_i W_i$$

where W_i represents an eigenvector and λ_i the corresponding eigenvalue. Since $\mu(M_a)$ is actually the solution to $\sup_U \rho(M_a U)$, there exists some supremizing $\hat{U}$ such that

$$\|M_a \hat{U} W_1\|_2 = \|\lambda_1(M_a \hat{U}) W_1\|_2 = \rho(M_a \hat{U})\|W_1\|_2 = \mu(M_a).$$

For $\|W_1\|_2 = 1$, this reduces to $\|M_a \hat{U} W_1\|_2 = \rho(M_a \hat{U}) = \mu(M_a)$. Defining vector $\hat{x} = \hat{U} W_1$ then gives

$$\|M_a \hat{x}\|_2 = \mu(M_a). \tag{7.1}$$

Computing the 2-norm of a complex 16×16 vector requires about $4n$ or 64 flops versus more than $50,000$ flops for the singular value decomposition of a complex 16×16

matrix. Therefore, optimizing the left-hand side of Equation 7.1 offers potentially significant numerical advantages in the calculation of μ.

The optimization must first be constrained to include only "x" vectors of the form UW_i. This constraint is imposed by the Fan and Tits formulation for $\mu(M_a)$:

$$\mu(M_a) = \max_{x \in C^{n^2}} \{\|M_a x\|_2 \mid \|\mathcal{P}_i x\|_2 \|M_a x\|_2 = \|\mathcal{P}_i M_a x\|_2,\ i = 1, \ldots, m\} \qquad (7.2)$$

where the $\mathcal{P}_i$'s are projection matrices with block size m and $x \in C^{n^2}$. Unfortunately, this function is nonconvex so it offers only candidate solutions. Determining whether these candidates equal $\mu(M_a)$ involves computing a corresponding scaling matrix D from x and ensuring that $\|M_a x\|_2 = \overline{\sigma}(DM_aD^{-1})$. Also, in its original formulation, this expression requires $n^2 - 1$ optimization variables. The following Theorem shows that the scaling structure introduced in Chapter 5 applies to Equation 7.2 so the Fan and Tits method actually requires no more than $2(n-1)$ optimization variables.

Theorem 7.1: Given $M \in C^{n \times n}$ and $\Delta \in C^{n \times n}$, where Δ is composed of 1×1 blocks, the solution for $\mu(M_a)$ in the constrained optimization problem

$$\mu(M_a) = \max_{x \in C^{n^2}} \{\|M_a x\|_2 \mid \|\mathcal{P}_i x\|_2 \|M_a x\|_2 = \|\mathcal{P}_i M_a x\|_2,\ i = 1, \ldots, m\}$$

with simple $\overline{\sigma}$ can be found with no more than $2(n-1)$ free variables.

Proof: As in Section 5.5, denote the reduced scaling structure of Equation 5.8 as S. Let $\hat{S}$ be the solution to $\inf_S \overline{\sigma}(SM_aS^{-1})$ and Y_1 the corresponding right singular vector. Then the optimal solution, $\hat{x}$, to Equation 7.2 may be written as [38]

$$\hat{x} = \frac{\hat{S}^{-1}Y_1}{\|\hat{S}^{-1}Y_1\|_2}.$$

The dependent relationships between the elements of Y_1 may be determined using the method of page 46. Since Y_1 is a singular vector of SM_aS^{-1}, any scalar multiple of Y_1 will also be a singular vector of SM_aS^{-1}. Therefore it can be scaled such that $y_{11} = 1$ without loss of generality. Likewise, $\hat{s}_1$ may also be scaled to 1. Carrying out the multiplication of $\hat{S}$ and Y_1, results in a general expression of the form

$$\hat{D}^{-1}Y_1 = \begin{bmatrix} 1 & \leftarrow fixed \\ y_{2,1} & \leftarrow free \\ \vdots & \vdots \\ y_{n,1} & \vdots \\ y_{n+1,1} & \leftarrow free \\ y_{2,1}y_{n+1,1} & \leftarrow fixed \\ \vdots & \vdots \\ y_{n,1}y_{n+1,1} & \leftarrow fixed \\ y_{2n+1,1} & \leftarrow free \\ y_{2,1}y_{2n+1,1} & \leftarrow fixed \\ \vdots & \vdots \\ y_{n,1}y_{(n-1)n+1,1} & \leftarrow fixed. \end{bmatrix} \tag{7.3}$$

Subtracting the $(n-1)^2$ fixed variables of Equation 7.3 from the $n^2 - 1$ free variables previously required leaves the desired result of $n^2 - 1 - (n-1)^2 = 2(n-1)$ free optimization variables. ■

Example 7.1 illustrates the application of Equation 7.3 to the system of Example 4.1.

Example 7.1: Using the system of Example 4.1, the optimal $\hat{S}^{-1}Y_1$ may be written as

$$S^{-1}Y_1 = \begin{bmatrix} \frac{y_{11}}{\hat{s}_1} \\ y_{11}\frac{\hat{s}_1 p_{12}}{\hat{s}_2^2 p_{11}} \\ y_{11}\frac{\hat{s}_1 p_{21}}{\hat{s}_3^2 p_{11}} \\ \frac{y_{41}}{\hat{s}_4} \\ y_{41}\frac{p_{12}}{\hat{s}_2^2 p_{11}\hat{s}_4} \\ y_{41}\frac{p_{21}}{\hat{s}_3^2 p_{11}\hat{s}_4} \\ \frac{y_{71}}{\hat{s}_7} \\ y_{71}\frac{p_{12}}{\hat{s}_2^2 p_{11}\hat{s}_7} \\ y_{71}\frac{p_{21}}{\hat{s}_3^2 p_{11}\hat{s}_7} \end{bmatrix}.$$

Scaling y_{11} and $\hat{s}_1$ to 1 gives

$$S^{-1}Y_1 = \begin{bmatrix} 1 \\ y_{21} \\ y_{31} \\ y_{41} \\ y_{21}y_{41} \\ y_{31}y_{41} \\ y_{71} \\ y_{21}y_{71} \\ y_{31}y_{71} \end{bmatrix} = \begin{bmatrix} 1 \\ 1.236 \\ 1.094 \\ 0.855 + 0.519\jmath \\ 1.056 + 0.641\jmath \\ 0.935 + 0.567\jmath \\ 0.355 + 0.935\jmath \\ 0.438 + 1.155\jmath \\ 0.388 + 1.023\jmath \end{bmatrix}$$

with corresponding

$$\|M_a\hat{x}\|_2 = 8.25 = \overline{\sigma}(DM_aD^{-1}).$$

7.2 Application of the Relationship Between R, L and S

As mentioned earlier, the solution to Equation 7.2 is nonconvex with many local maxima possible. The only means of ensuring that a candidate solution vector, x,

actually corresponds to $\mu(M_a)$ is to determine optimal scaling matrix $\hat{D}$ such that

$$\|M_a x\|_2 = \overline{\sigma}(\hat{D} M_a \hat{D}^{-1}) = \mu(M_a). \tag{7.4}$$

(The method for directly computing D given x is presented in Section 8.6.) Unfortunately, this singular value decomposition must be performed on the expanded $n^2 \times n^2$ system for each candidate solution, reducing the overall efficiency of this technique. However, using the direct transformation from D to R and L presented in Chapter 6, the verification test for $\mu(M_a)$ may be reformulated as

$$\|M_a x\|_2 = \overline{\sigma}(R^{-1} M L^{-1}) \overline{\sigma}(LPR) \tag{7.5}$$

where the left-hand side consists of only $n \times n$ matrices. For a system with $n = 3$, the expanded $n^2 \times n^2$ matrices of Equation 7.4 require about $10,000$ flops for a single singular value decomposition. Equation 7.5, on the other hand, requires about $1,000$ flops including both singular value decompositions and the required transformation from D to L and R. Even more substantial savings occur as n increases.

The Fan and Tits method outlined in this chapter first appeared in 1985 [37]. Their constrained optimization procedure was implemented using specialized software from the proprietary Harwell Subroutine Library [39]. Since that time, the growing acceptance of the structured singular value concept has led to a number of alternative formulations of μ. These formulations are aimed at improving its computational properties to allow for interactive design without the restrictions imposed by the use of proprietary software. The recent announcement by the MathWorks of the "μ-Analysis and Synthesis Toolbox" for MATLAB^{TM} indicates the demand for

user-friendly software to calculate μ [40]. This new toolbox uses a variation of the power method proposed by Packard, Fan, and Doyle to compute a lower bound for $\mu(M_a)$ [41]. This optimization searches for stationary points of the singular values of (DM_aD^{-1}) with respect to the elements of D. From the Principal Direction Alignment theory of Daniel, Kouvaritakis, and Latchman, such stationary points have been shown to coincide with stationary points of $\rho(M_aU_d)$ with respect to the elements of U_d [31]. However, like all known lower bound expressions, the procedure is nonconvex so candidate solutions for $\mu(M_a)$ must be compared to the singular value upper bound to see if they actually achieve the true supremum. Again, rather than performing a singular value decomposition on the $n^2 \times n^2$ system DM_aD^{-1}, the test may be formulated as

$$\tilde{\mu}(M_a) \leq \overline{\sigma}(R^{-1}ML^{-1})\overline{\sigma}(LPR) \tag{7.6}$$

involving ly $n \times n$ matrices where $\tilde{\mu}(M_a)$ represents a candidate for $\mu(M_a)$. If the inequality of Equation 7.6 holds with equality, then $\tilde{\mu}(M_a) = \mu(M_a)$.

If no candidate for μ is found for which the lower and upper bounds become equal, then the system may have a repeated maximum singular value at the "inf" of Equation 5.1 for scalar uncertainties or Equation 5.9 for block structured uncertainties. Lack of stationarity at this upper bound implies that no $\tilde{\mu}$ exists for which the upper and lower bounds become equal. Chapter 8 contains a detailed description of the difficulties involved with computing μ when the maximum singular value repeats.

CHAPTER 8
CALCULATION OF THE STRUCTURED SINGULAR VALUE WITH REPEATED MAXIMUM SINGULAR VALUES

8.1 Cusping Singular Values

The work presented in the previous chapters generally requires that the maximum singular value not repeat. This requirement stems from the fact that the inequality in the right-hand side of

$$\mu(M_a) = \sup_{U_d} \rho(M_a U_d) \leq \inf_{D} \overline{\sigma}(D M_a D^{-1})$$

is not guaranteed to achieve equality at the "inf" if $\overline{\sigma} = \sigma_i$, $i \neq 1$ because this suggests that a stationary point with respect to D has not been found. Following the infimization of $\overline{\sigma}(DM_aD^{-1})$ with respect to D, three possible conditions can occur. These three conditions are illustrated in Figures 8.1 through 8.3 with corresponding M_a matrices in Tables 8.1 through 8.3 respectively. Figure 8.1 represents a system in which $\inf_D \overline{\sigma}(DM_aD^{-1})$ is achieved with distinct $\overline{\sigma}(DM_aD^{-1})$. From the MPDA property reviewed in Chapter 4, this guarantees the simultaneous determination of $\sup_{U_d} \rho(M_aU_d)$ and hence $\mu(M_a)$.

Figure 8.2 represents a second system where $\inf_D \overline{\sigma}(DM_aD^{-1})$ occurs with $\overline{\sigma} = \sigma_2 = \sup_{U_d} \rho(M_aU_d)$. The point at which the equalities hold will be referred to as a "kiss" and its calculation requires special consideration.

Figure 8.3 shows a system in which $\inf_D \overline{\sigma}(DM_aD^{-1})$ occurs with $\overline{\sigma} = \sigma_2 \neq \sup_{U_d} \rho(M_aU_d) = \mu(M_a)$. This case will be referred to as a "cusp" and

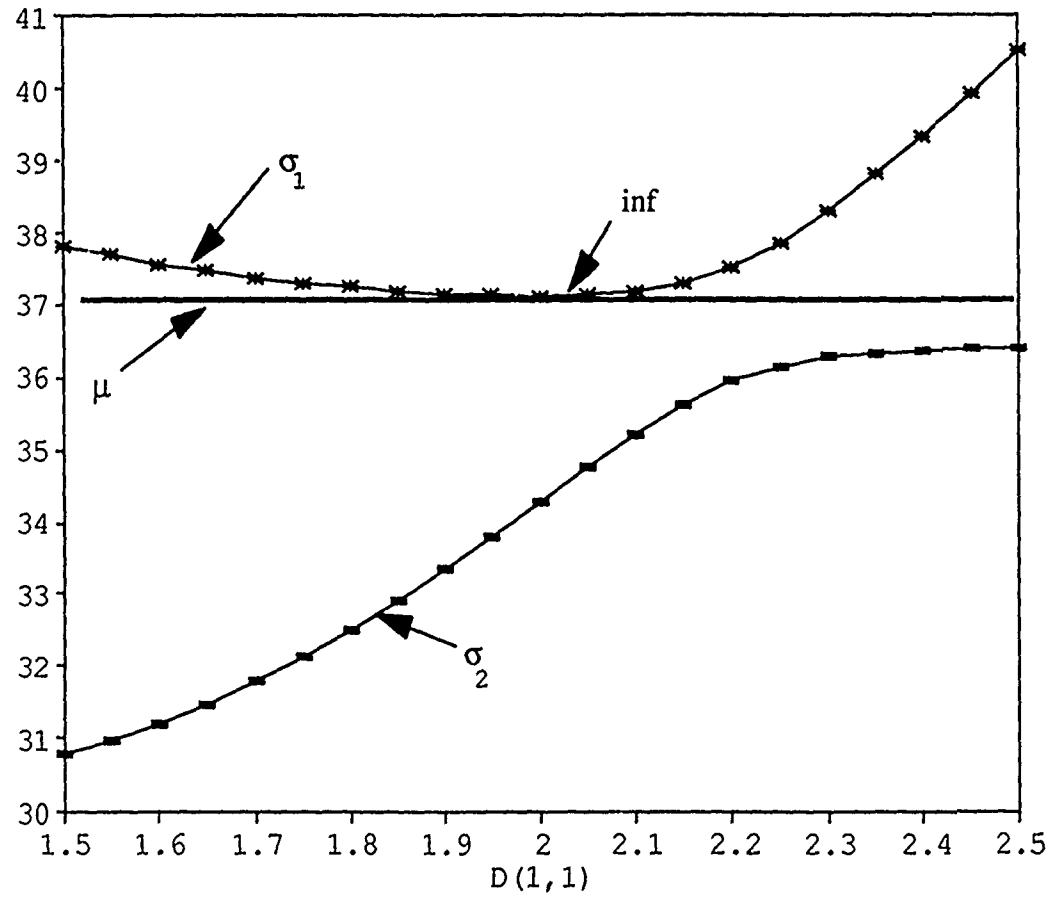

Figure 8.1: Noncusping System.

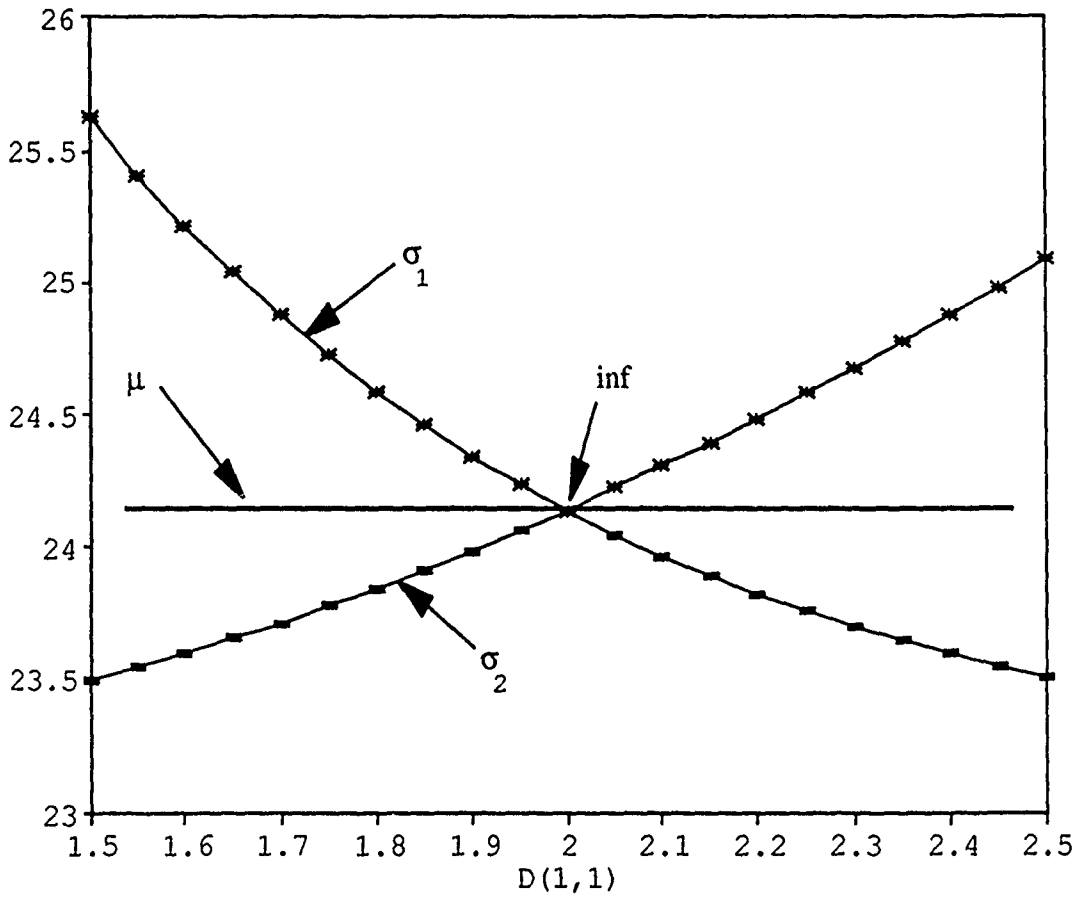

Figure 8.2: System with "Kiss".

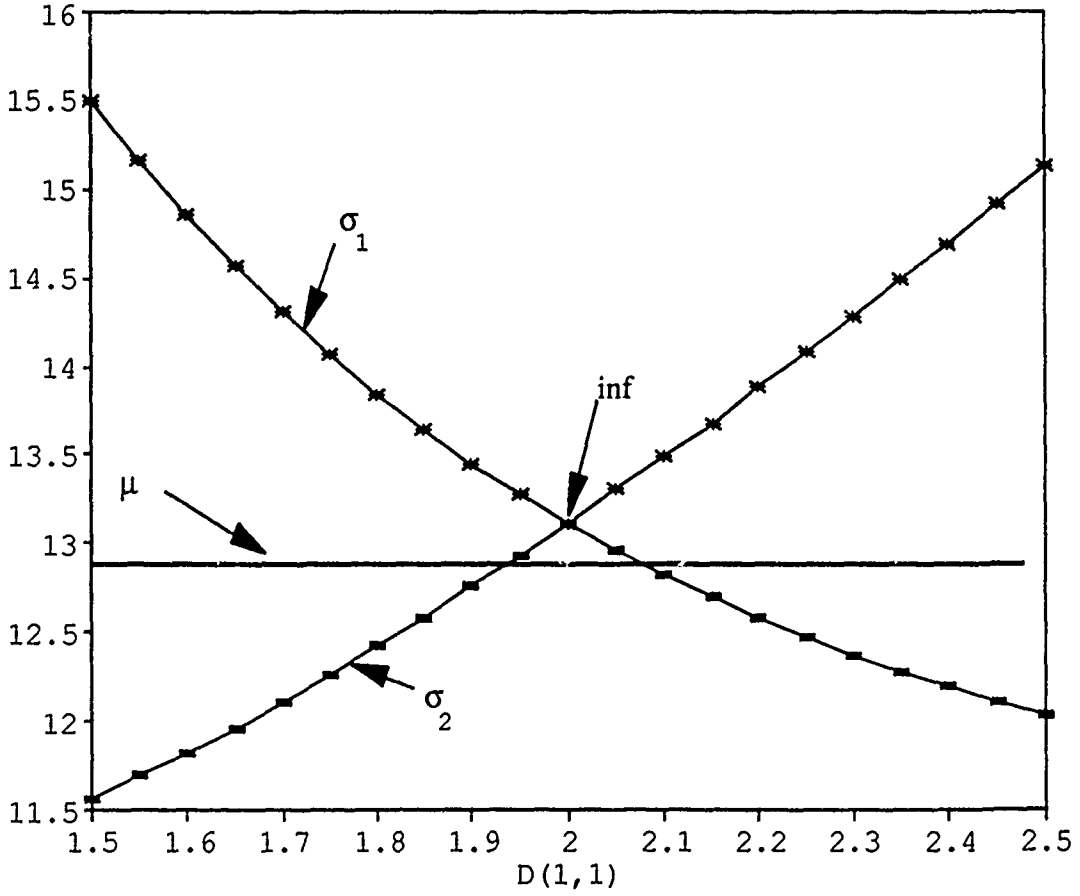

Figure 8.3: System with "Cusp".

Table 8.1: M_a Matrix for Figure 8.1.

$$\begin{bmatrix} -0.89+2.04j & -4.34-5.31j & -1.83+4.69j & 3.49-6.31j & 3.07+5.92j \\ 3.87+17.96j & 1.66-4.72j & 6.83+2.94j & -7.38-14.37j & -8.73-1.29j \\ -7.68-3.20j & -11.10-8.20j & 6.84-20.10j & -3.15+0.62j & 2.27+7.21j \\ -5.11-11.27j & -2.81+14.23j & 8.19+11.80j & 8.89+14.60j & 2.73-10.69j \\ 9.83-18.68j & -5.39+6.77j & 1.11+0.19j & 6.29+2.55j & -2.83+6.75j \end{bmatrix}$$

Table 8.2: M_a Matrix for Figure 8.2.

$$\begin{bmatrix} 19.18+0.37j & 6.82-1.75j & 3.13-0.95j & -4.92+1.11j & 3.34-4.59j \\ -0.20-3.07j & 18.56+1.29j & -1.44+0.35j & 3.22+2.37j & -1.32+3.15j \\ 6.42+1.85j & -0.70+1.03j & 15.34-2.01j & -0.77-0.82j & -0.13+1.36j \\ 0.06+0.64j & -0.53-2.47j & 3.53-0.97j & 10.96-3.11j & 3.93-0.96j \\ -2.39-5.34j & 3.21-0.78j & 3.74+1.38j & 3.24-0.03j & 15.41-0.33j \end{bmatrix}$$

Table 8.3: M_a Matrix for Figure 8.3.

$$\begin{bmatrix} 5.18+0.37j & 6.82-1.75j & 3.13-0.95j & -4.92+1.11j & 3.34-4.59j \\ -0.20-3.07j & 4.56+1.29j & -1.44+0.35j & 3.22+2.37j & -1.32+3.15j \\ 6.42+1.85j & -0.70+1.03j & 1.34-2.01j & -0.77-0.82j & -0.13+1.36j \\ 0.06+0.64j & -0.53-2.47j & 3.53-0.97j & -3.03-3.11j & 3.93-0.96j \\ -2.39-5.34j & 3.21-0.78j & 3.74+1.38j & 3.24-0.03j & 1.41-0.33j \end{bmatrix}$$

calculating $\mu(M_a)$ for such a system represents a quite challenging problem. The remainder of this chapter will address several approaches that greatly improve upon previous techniques for determining $\mu(M_a)$ when $\overline{\sigma}(DM_aD^{-1})$ repeats.

As previously discussed, noncusping systems such as that of Figure 8.1 are guaranteed to achieve μ at $\inf_D \overline{\sigma}(DM_aD^{-1})$. Using the following theorem due to Latchman [27], it is possible to compute the supremizing U_d of $\sup_{U_d} \rho(M_aU_d)$ given the infimizing D of $\inf_D \overline{\sigma}(DM_aD^{-1})$.

Theorem 8.1: Given the optimal diagonal scaling matrix $\hat{D}$ corresponding to a stationary point, there exists a diagonal unitary matrix, $\hat{U}_d = diag[e^{j\theta_1}, \ldots, e^{j\theta_n}]$, $0 \leq \theta_i \leq 2\pi$, where $\theta_i = [\arg(y_{i1}) - \arg(x_{i1})]$, $i = 1, \ldots, n$, such that the principal singular vectors X_1 and Y_1 of $\hat{D}M_a\hat{D}^{-1}\hat{U}_d$ are aligned.

Proof: Recall that $\rho(M_aU_d) \leq \overline{\sigma}(DM_aD^{-1}U_d)$. For any system, singular vectors X_1 and Y_1 obey the following relationships:

$$\begin{aligned} (\hat{D}M_a\hat{D}^{-1})Y_1 &= \overline{\sigma}X_1 \\ X_1^H(\hat{D}M_a\hat{D}^{-1}) &= \overline{\sigma}Y_1^H. \end{aligned} \tag{8.1}$$

Introducing a diagonal unitary matrix, U_d, allows Equations 8.1 to be rewritten as

$$\begin{aligned} (\hat{D}M_a\hat{D}^{-1}U_d)U_d^HY_1 &= \overline{\sigma}X_1 \\ X_1^H(\hat{D}M_a\hat{D}^{-1}U_d) &= \overline{\sigma}Y_1^HU_d. \end{aligned} \tag{8.2}$$

From Equations 8.1 and 8.2 it is apparent that U_d may alter the phases of X_1 and Y_1 while continuing to satisfy the MPDA requirement that the magnitudes of the corresponding elements of X_1 and Y_1 be equal. The independent phase alignment

of the elements of X_1 and Y_1 provides the necessary information to compute the supremizing U_d (denoted as $\hat{U}_d$). ■

Using Theorem 8.1, the individual diagonal elements of $\hat{U}_d$ are found to be

$$\hat{u}_i = e^{\jmath[\arg(y_{i1})-\arg(x_{i1})]} = e^{\jmath\hat{\theta}_i}, \quad i = 1, \ldots, n \tag{8.3}$$

for element-by-element structured uncertainties. A similar expression reflecting individual uncertainty blocks applies to the general block structured case. The actual $\hat{U}_d$ determined in this manner is unique within some scalar multiplier $e^{\jmath\phi}$.

Example 8.1: Applying the relationship of Equation 8.3 with

$\hat{D} = diag[1, 0.460, 0.350, 0.425, 0.541]$, the corresponding $\hat{U}_d$ for the system of Figure 8.1 is

$$\hat{U}_d = diag[1,\ 0.996 - \jmath 0.078,\ -0.998 + \jmath 0.060,\ 0.966 + \jmath 0.250,\ 0.921 - \jmath 0.390]$$

with $\overline{\sigma}(\hat{D}M_a\hat{D}^{-1}) = \rho(M_a\hat{U}_d) = 36.94 = \mu(M_a)$.

The ability to directly compute $\hat{U}_d$ given $\hat{D}$ provides a simple means to ensure that $\overline{\sigma}(\hat{D}M_a\hat{D}^{-1}) = \mu(M_a)$ because it guarantees that $\sup_{U_d} \rho(M_aU_d) = \inf_D \overline{\sigma}(DM_aD^{-1})$. This becomes increasingly important as the conditions illustrated in Figure 8.2 are considered.

8.2 Systems with $\inf_D \overline{\sigma}(DM_aD^{-1}) = \sigma_2(DM_aD^{-1}) = \mu(M_a)$

When the singular values become equal or nearly equal, the corresponding singular vectors become unstable with small changes in the σ_i's resulting in large variations in the columns of X and Y [42]. This behavior is a direct result of the

expressions defining the singular vectors of a matrix A, namely

$$\begin{aligned} AA^H X_i &= \sigma_i^2 X_i \\ A^H A Y_i &= \sigma_i^2 Y_i. \end{aligned} \tag{8.4}$$

When the singular values repeat, the singular vectors combine to span a $q-$dimensional subspace where q represents the multiplicity of the repeating $\sigma_{i's}$. For $q = 2$, (the most common case), the first line of Equation 8.4 becomes

$$\begin{aligned} AA^H X_1 &= \sigma_1^2 X_1 \\ AA^H X_2 &= \sigma_1^2 X_2 \end{aligned} \tag{8.5}$$

with the resulting subspace formed by the linear combination of X_1 and X_2 to give

$$AA^H[(\gamma + \jmath\delta)X_1 + (\alpha + \jmath\beta)X_2] = \sigma_1^2[(\gamma + \jmath\delta)X_1 + (\alpha + \jmath\beta)X_2]. \tag{8.6}$$

(Note that Y_i is dependent on X_i through the relationship $A^H X_i = \sigma_i Y_i$ and thus a similar development for the subspaces spanned by Y_1 and Y_2 would be redundant). From Equation 8.6 it is apparent that any linear combination of X_1 and X_2 satisfies the definition of a singular vector. This arbitrariness poses a significant difficulty for algorithms relying on the vector alignment properties of MPDA theory because there is no longer any guarantee that vectors chosen arbitrarily from the $q-$dimensional subspace will satisfy the alignment requirements.

For example, the system of Figure 8.2 has $\inf_D \overline{\sigma}(DM_aD^{-1}) = 24.14$. Using the method of Theorem 8.1 for X_1, Y_1 and X_2, Y_2 gives corresponding values of $\rho(M_aU_d)$ as follows:

$$X_1, Y_1 \rightarrow \rho(M_a U_d) = 23.12 \neq \mu(M_a)$$

$$X_2, Y_2 \rightarrow \rho(M_a U_d) = 20.70 \neq \mu(M_a)$$

Both the first and second principal directions must be considered in this case since both represent legitimate solutions to Equation 8.5. The actual vectors from the $q-$dimensional subspace that do provide the supremizing $\hat{U}_d$ may be found by selecting a vector X_{new} defined as

$$X_{new} = (\gamma + \jmath\delta)X_1 + (\alpha + \jmath\beta)X_2$$

as in Equation 8.6. Of course X_{new} is a singular vector of $(\hat{D}M_a\hat{D}^{-1})$ and therefore any scalar multiple of X_{new} is also a singular vector of $(\hat{D}M_a\hat{D}^{-1})$. This allows the complex multiplier $(\gamma + \jmath\delta)$ of X_1 to be eliminated with no loss of generality leaving only α and β as unknowns. Vector X_{new} and the corresponding vector Y_{new} may now be written as

$$\begin{aligned} X_{new} &= X_1 + (\alpha + \jmath\beta)X_2 \\ Y_{new} &= \frac{(\hat{D}M_a\hat{D}^{-1})^H X_{new}}{\overline{\sigma}} \end{aligned} \tag{8.7}$$

and a 2−dimensional optimization performed over α and β to achieve the MPDA conditions of aligning $|X_{new}|$ and $|Y_{new}|$. Assigning the complex multiplier $(\alpha + \jmath\beta)$ to X_2 with initial conditions of $\alpha = \beta = 0$ allows this procedure to be implemented very efficiently even when $\overline{\sigma} \neq \sigma_2$. This assignment serves to initially weight X_1 more heavily for cases with $\overline{\sigma} > \sigma_2$ while still providing for subspace combinations if $\overline{\sigma}$ and σ_2 coalesce.

Corollary 8.1: For systems with repeated maximum singular value of multiplicity q with

$$\inf_D \overline{\sigma}(DM_aD^{-1}) = \sigma_2(DM_aD^{-1}) = \cdots = \sigma_q(DM_aD^{-1}) = \mu(M_a),$$

the following condition must hold:

$$\mu(M_a) = \rho(M_aU_d) \mid |X_{new}(\alpha_i)| - |Y_{new}(\alpha_i)| = 0 \quad i = 1, \ldots, 2(q-1). \tag{8.8}$$

Proof: From the statement of the Corollary it must be possible to establish MPDA because $\overline{\sigma}(DM_aD^{-1}) \neq \mu(M_a)$ otherwise. This requires that some linear combination of the first q columns of X and Y respectively must attain alignment. When alignment is achieved, the equality of Equation 8.8 must hold. ■

The resulting $q-$dimensional optimization requires only vector operations to align X_{new} and Y_{new}. Although this function is not convex, it requires only $2(q-1)$ optimization variables regardless of the size of the system and thus proceeds very quickly.

Example 8.2: Using the system of Figure and Table 8.2, the solution to Equation 8.8 occurs at $\alpha = -0.0009$, and $\beta = 0.0921$. The corresponding optimal $\hat{U}_d$ is

$$\hat{U}_d = diag[1,\ 0.91 + \jmath 0.43,\ 1.0 + \jmath 0.01,\ -.94 - \jmath 0.33,\ 0.41 + \jmath 0.91]$$

with $\rho(M_a\hat{U}_d) = \overline{\sigma}(\hat{D}M_a\hat{D}^{-1}) = \mu(M_a) = 24.14$.

The values of α and β are not robust as small changes in $\hat{D}$ cause the singular vectors to range throughout the $2-$dimensional subspace spanned by X_1 and X_2. This is important considering that algorithms used to solve $\inf_D \overline{\sigma}(DM_aD^{-1})$ depend on the analytic derivative, $\frac{\partial\overline{\sigma}}{\partial d_i}$, as developed in Section 5.5. Since the gradient is a function

of X_1 and Y_1, as $\overline{\sigma}$ approaches σ_2 the singular vectors, and hence the gradient, begin to change discontinuously. This makes it extremely difficult to determine whether a candidate solution for μ using only the gradient at $\inf_D \overline{\sigma}(DM_aD^{-1})$ is a true cusp or simply a kiss. Through the use of Equation 8.8 and the method of Theorem 8.1 however, the system of Figure 8.2 is shown to be a kiss.

8.3 Systems with $\inf_D \overline{\sigma}(DM_aD^{-1}) = \sigma_2(DM_aD^{-1}) \neq \mu(M_a)$

For the case of a true cusp as depicted in Figure 8.3, it is not possible to rotate $|X_{new}|$ into $|Y_{new}|$ because this condition only occurs if MPDA can be attained. Although no theoretical bound on the gap between $\inf_D \overline{\sigma}(DM_aD^{-1})$ and $\mu(M_a)$ has been determined, experimental tests on cusping systems have revealed disagreements of more than 14% [43]. Determining the actual value of μ previously required solving for $\sup_{U_d} \rho(M_aU_d)$ over $n-1$ free variables where n denotes the size of expanded matrix M_a. Since this computationally intensive optimization is nonconvex the lower bound by itself is not very useful since it is difficult to determine which of the local maxima is actually μ. Therefore, the upper bound of $\overline{\sigma}(\hat{D}M_a\hat{D}^{-1})$ is generally chosen as an estimate for $\mu(M_a)$ even though this leads to an unnecessarily conservative control system .

Using a method similar to that of the previous section, it may still be possible to find an X_{new} that produces a U_d (denoted U_{new}) "close" to the supremizing $\hat{U}_d$. This possibility arises from the fact that while vectors $X_1, \ldots, X_q$ may individually change greatly with small variations in $(\hat{D}M_a\hat{D}^{-1})$, the actual subspace spanned by linear combinations of $X_1, \ldots, X_q$ remains continuous [44, 45]. Combining this

important property with the following lemma leads to an extremely efficient method for approximating $\mu(M_a)$ for cusping systems.

Lemma 8.1: For any $A \in C^{n\times n}$ composed of elements a_{ij}, $i,j = 1,\ldots,n$,

$$\lim_{a_{ii}\to\pm\infty} \overline{\sigma}(A) = |a_{ii}| \quad \text{for fixed } i. \tag{8.9}$$

Proof: The singular values of A are the square roots of the eigenvalues of A^HA. As an element along the diagonal of A increases or decreases towards $\pm\infty$, the remaining elements of A become insignificant in comparison. At this point matrix product A^HA may be approximated by $\overline{a_{ii}}a_{ii}E_{ii}$ making the maximum singular value equal to $|a_{ii}|$. ■

Lemma 8.1 simply shows that by perturbing a diagonal element of a cusping system, the multiplicity of $\overline{\sigma}$ must eventually disappear. In fact, combinations of the diagonal elements may be perturbed providing a number of different ways to break up a cusp. This shift eliminates the cusp and allows the establishment of MPDA for the shifted matrix (denoted $\tilde{M}_a$) guaranteeing that $\inf_D \overline{\sigma}(D\tilde{M}_aD^{-1}) = \mu(\tilde{M}_a)$. MPDA conditions also provide a direct method for computing $\hat{U}_d$ given $\hat{D}$ through the use of Theorem 8.1. Modifying this method to use X_{new} and Y_{new} from Equation 8.7 leads to the following Proposition.

Proposition 8.1: Given some $M_a \in C^{n\times n}$ for which

$$\inf_D \overline{\sigma}(DM_aD^{-1}) = \sigma_2(DM_aD^{-1}) = \cdots = \sigma_q(DM_aD^{-1}) \neq \mu(M_a),$$

the solution over only $2(q-1)$ free variables of

$$\sup_{\alpha_i} \rho(M_a U_{new}) \quad i = 1, \ldots, 2(q-1) \tag{8.10}$$

is a close approximation of $\mu(M_a)$.

Discussion: The continuity of the spanning subspace $X_1 \cdots X_q$ implies that small perturbations in the matrix elements of M_a will result in correspondingly small perturbations in the subspace. As a scalar indication of the distance between two matrices $A, B \in C^{n \times n}$, define an error function $E = \|A - B\|_2$ [42]. Shifting the diagonal elements of the cusping system in Table 8.3 until the cusp disappears produces a new system with desirable MPDA properties. If this new system is sufficiently close to the original cusping system then X_{new} of the two systems should be similar.

Examination of Tables 8.2 and 8.3 reveals that the two systems are identical except for an additive shift of +14 to all the diagonal elements of Table 8.2 so that $\tilde{M}_a = M_a + 14 \cdot I_5$. Denote M_a and the infimizing D of Table 8.2 as M_{a_2} and $\hat{D}_2$ respectively. Similarly denote the corresponding values of Table 8.3 as M_{a_3} and $\hat{D}_3$. Then

$$E = \|\hat{D}_3 M_{a_3} \hat{D}_3^{-1} - \hat{D}_2 M_{a_2} \hat{D}_2^{-1}\| = 16.44.$$

Since $\overline{\sigma}(\hat{D}_3 M_{a_3} \hat{D}_3^{-1}) = 13.11$ from Figure 8.3, E indicates that the two systems are not particularly close.

However, an alternative shift of only element $M_a(1,1)$ by +4 also breaks up the cusp. Figure 8.4 and Table 8.4 show this new system with

$$\mu(M_a) = \overline{\sigma}(\hat{D}_4 M_{a_4} \hat{D}_4^{-1}) = 15.182.$$

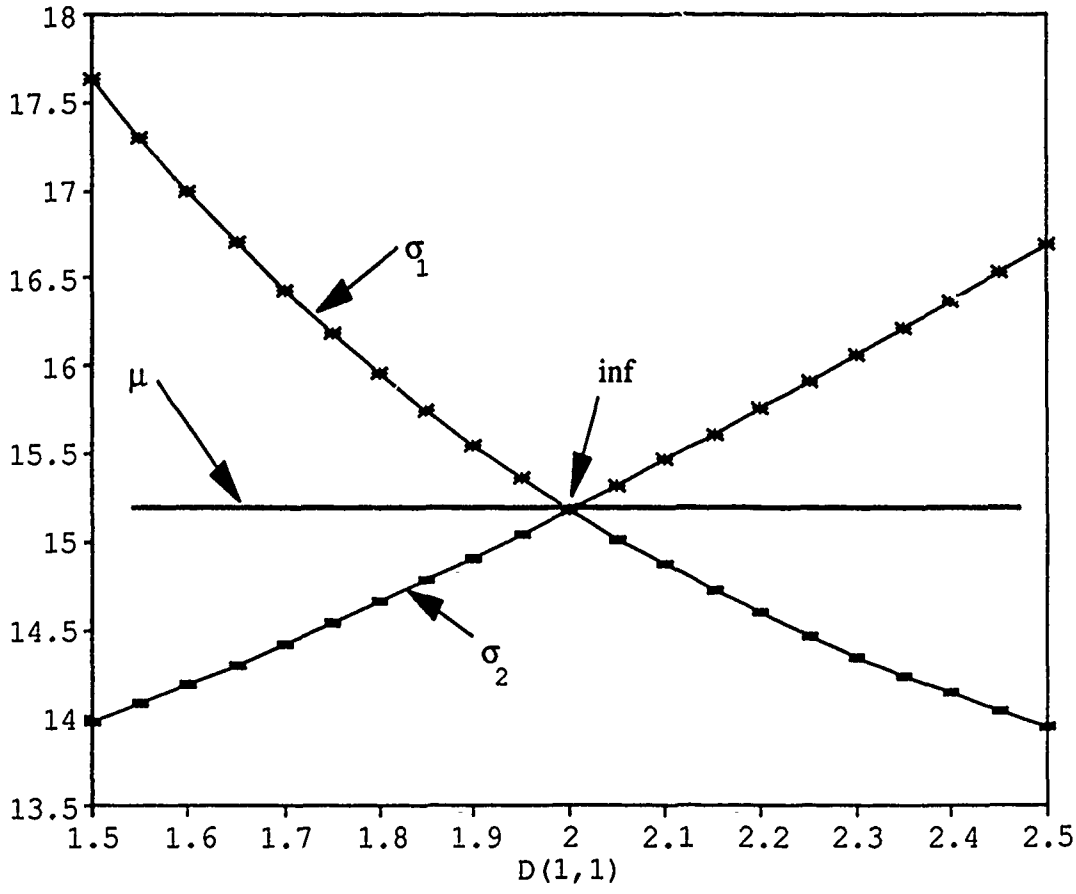

Figure 8.4: System from Figure 8.3 with Shifted $M_a(1,1)$ Element.

Table 8.4: M_a Matrix for Figure 8.4.

$$\begin{bmatrix} 9.18+0.37j & 6.82-1.75j & 3.13-0.95j & -4.92+1.11j & 3.34-4.59j \\ -0.20-3.07j & 4.56+1.29j & -1.44+0.35j & 3.22+2.37j & -1.32+3.15j \\ 6.42+1.85j & -0.70+1.03j & 1.34-2.01j & -0.77-0.82j & -0.13+1.36j \\ 0.06+0.64j & -0.53-2.47j & 3.53-0.97j & -3.03-3.11j & 3.93-0.96j \\ -2.39-5.34j & 3.21-0.78j & 3.74+1.38j & 3.24-0.03j & 1.41-0.33j \end{bmatrix}$$

Error function E then becomes

$$E = \|\hat{D}_3 M_{a_3} \hat{D}_3^{-1} - \hat{D}_4 M_{a_4} \hat{D}_4^{-1}\| = 4.77$$

indicating that System 4 is much closer to System 3 than System 2 was. Since the spanning subspace is continuous, it seems reasonable for the optimum $\hat{X}_{new} = X_1 + (\hat{\alpha}_1 + \jmath\hat{\alpha}_2)X_2$ of the cusping system to be similar to that of Table 8.4 for which MPDA is established. Table 8.5 shows the variation of $\hat{X}_{new}$ as the cusping system of Table 8.3 (Shift=0) is shifted to the noncusping system of Table 8.4 (Shift=4).

Table 8.5: Variation of $\hat{X}_{new}$ due to Shift of Table 8.3 System.

Shift of $M_a(1,1)$				
+0	+1	+2	+3	+4
1	1	1	1	1
-0.053- $\jmath$ 0.237	-0.070 - $\jmath$ 0.248	-0.080 - $\jmath$ 0.254	-0.089 - $\jmath$ 0.258	-0.10 - $\jmath$ 0.26
0.425+ $\jmath$ 0.210	0.424 + $\jmath$ 0.166	0.409 + $\jmath$ 0.135	0.388 + $\jmath$ 0.112	0.37 + $\jmath$ 0.09
0.300+ $\jmath$ 0.240	0.309 + $\jmath$ 0.246	0.314 + $\jmath$ 0.244	0.315 + $\jmath$ 0.241	0.31 + $\jmath$ 0.24
-0.027- $\jmath$ 0.345	-0.058 - $\jmath$ 0.352	-0.075 - $\jmath$ 0.355	-0.087 - $\jmath$ 0.354	-0.09 - $\jmath$ 0.35

Since shifting other combinations of matrix elements will also break up the cusp, it may certainly be possible to find noncusping systems for which E is less than the 4.77 achieved by System 4. Regardless of the shifting combination used to break up a cusp, the $2(q-1)$-dimensional subspace spanned by $X_1, \ldots, X_q$ must vary continuously from the cusping condition to any of the noncusping conditions. This implies that the spanning subspace is sufficiently well-behaved to provide an $\hat{X}_{new}$ from Equation 8.7 and corresponding $\hat{U}_{new}$ that closely approximates $\hat{U}_d$ of $\rho(M_a\hat{U}_d) = \mu(M_a)$. Although no bounds have been placed on the difference between $\sup_{\alpha_i} \rho(M_a U_{new})$ and $\sup_{U_d} \rho(M_a U_d)$, numerical experience suggests that this difference is quite small. In

fact, the largest deviation between this estimate and μ found to date is less than 0.34%. ■

Applying Equation 8.10 to the cusping system of Table 8.3 gives

$$\hat{U}_{new} = diag[1,\ 0.75 - \jmath 0.67,\ -.48 + \jmath 0.88,\ -1.00 - \jmath 0.02,\ 0.78 - \jmath 0.63]$$

with $\alpha_1 = -.23$, $\alpha_2 = -0.11$ and $\rho(M_a\hat{U}_{new}) = 12.768$.

To determine whether U_{new} is actually a candidate for μ, the derivative $\partial\rho/\partial\theta_i$ for $i = 1, \ldots, n$ at $\hat{U}_{new}$ may be examined for its proximity to the 0 vector. In this case it is found to be

$$\frac{\partial\rho}{\partial\theta_i} = \begin{bmatrix} 0.24 \\ 0.10 \\ -0.19 \\ 0.31 \\ 0.17 \end{bmatrix}$$

indicating that $\hat{U}_{new}$ is indeed "almost" a stationary point. Using $\hat{U}_{new}$ as an initial condition, a direct supremization over all U_d returns a value of

$$\hat{U}_d = diag[1,\ 0.59 - \jmath 0.80,\ -.53 + \jmath 0.85,\ -.99 - \jmath 0.11,\ 0.66 - \jmath 0.75] \qquad (8.11)$$

with $\rho(M_a\hat{U}_d) = \mu(M_a) = 12.81$. The difference between $\rho(M_a\hat{U}_{new})$ and $\rho(M_a\hat{U}_d)$ in this case is 0.33%.

This particular example was chosen because it produced the largest percent difference between $\rho(M_a\hat{U}_{new})$ and $\mu(M_a)$ out of over 100 cusping systems tested. The next largest percent difference was 10 times smaller than this case: about 0.03%.

While larger differences are certainly possible, the fact that in the majority of examples investigated, $\rho(M_a\hat{U}_{new})$ and $\rho(M_a\hat{U}_d)$ agreed within four significant figures is most encouraging. Table 8.6 on page 114 contains a comparison between $\mu(M_a)$ and $\rho(M_a\hat{U}_{new})$ for several cusping systems. (This comparison will be discussed more completely in Section 8.8.)

8.4 Effects of Nonconvexity

It must be emphasized that both the $2(q-1)$-dimensional search over the α_i and the $n-1$ dimensional search over U_d are nonconvex. Algorithms designed to find the global maximum must employ some procedure to search the domains of U_d or U_{new} respectively. Therefore, the advantages of the $2(q-1)$-dimensional search become more profound because the domain of U_{new} does not increase with system size.

Although repeated singular values with multiplicity $q = n$ are theoretically possible, by far the more common case is for $q = 2$. For such systems, an additional bonus of a $2(q-1) = 2-$dimensional search is that $3-$dimensional mesh plots of $\rho(M_aU_d)$ vs α_1 and α_2 may be generated allowing all maxima to be examined. This effectively eliminates the difficulties of multiple maxima because all candidate points for μ can easily be examined.

Such a mesh plot for the system of Figure 8.3 is presented in Figure 8.5. In this plot, α_1 and α_2 range from -2 to 2. Two pronounced local maxima are readily apparent: one corresponding to the value of $\rho(M_a\hat{U}_{new}) = 12.768$ and the other at

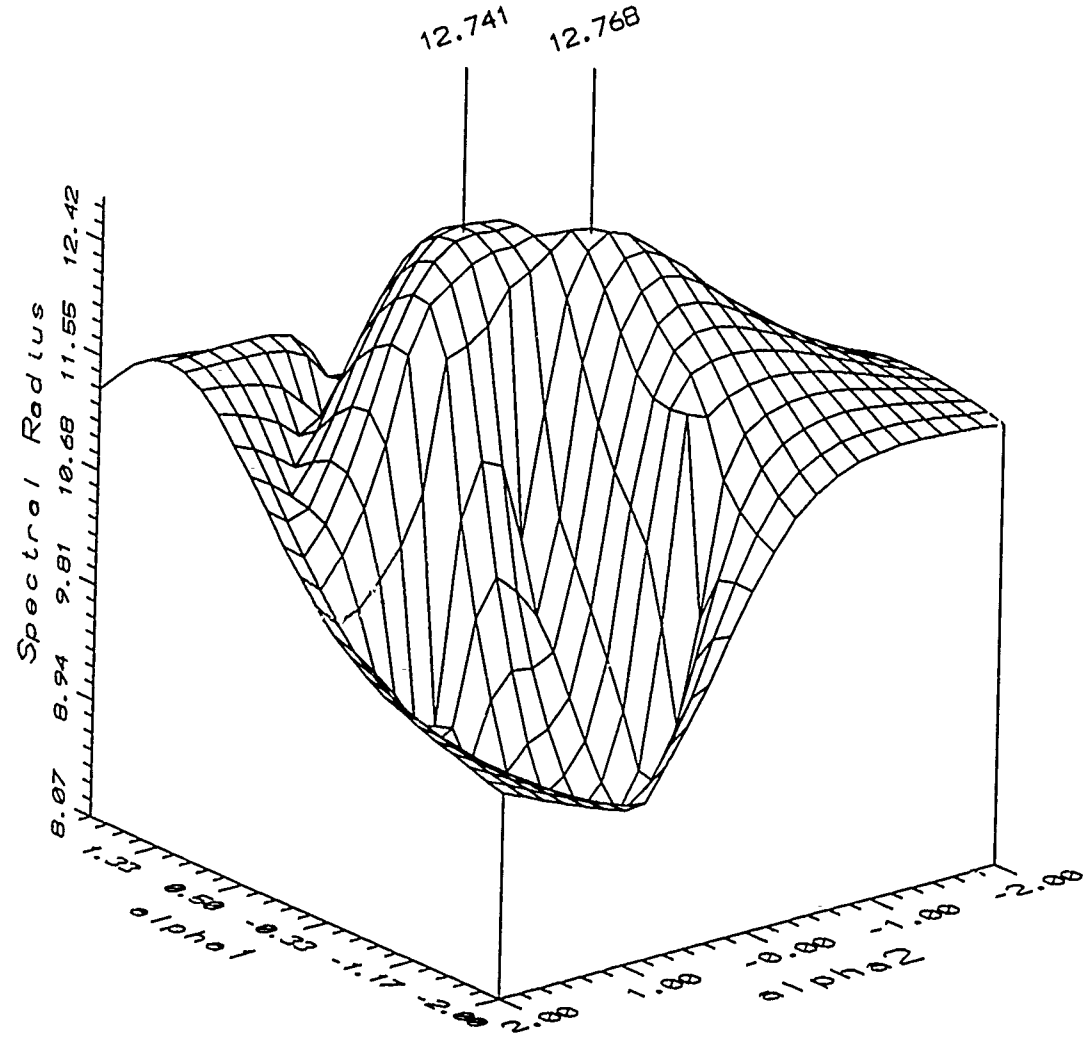

Figure 8.5: Mesh Plot of System from Table 8.3.

$\rho(M_a U_{new}) = 12.741$ where

$$U_{new} = diag[1,\ -.13 + \jmath 0.99,\ 1.00 + \jmath 0.084,\ -.91 + \jmath 0.42,\ -0.81 + \jmath 0.58]$$

with $\alpha_1 = 0.25$ and $\alpha_2 = 0.54$. Using U_{new} as a starting point, a search over all U_d results in a local stationary point at

$$U_d = diag[1,\ -.17 + \jmath 0.98,\ 0.95 + \jmath 0.32,\ -.94 + \jmath 0.33,\ -.92 + \jmath 0.40]$$

with corresponding $\rho(M_a U_d) = 12.76 \neq \mu(M_a) = 12.81$.

8.5 Range of α_i

The ability to easily select local maxima of $\rho(M_a U_{new})$ from a mesh plot as candidates for μ is a tremendous improvement over previous methods which require optimizing $\rho(M_a U_d)$ over as many as $n - 1$ variables, especially as the system size increases. The range of α_1 and α_2 used in Figure 8.4 appears to adequately cover the subspace spanned by X_1 and X_2 allowing $\hat{U}_{new}$ to be found. As α_1 and α_2 move away from the local maxima, the magnitude of ρ decreases suggesting, (but not guaranteeing), that no additional points of interest exist outside of the plotted region. It is important to determine the maximum range of the α_i because if the bounds can only be placed at $\pm\infty$, then a mesh plot or any other search method would clearly be impractical.

For the case of multiplicity $q = 2$, determining limits on α_1 and α_2 requires examination of Equation 8.7. Placing the complex multiplier $(\alpha_1 + \jmath\alpha_2)$ on X_2 makes the worst case condition when the optimum $\hat{X}_{new} = X_2$ so that X_1 makes no contribution. Such a condition implies that $\alpha_1 = \infty$ is required to produce the desired

$\hat{X}_{new} = X_2$. However, since X_1 and X_2 are singular vectors they may be normalized such that $\|X_1\| = \|X_2\| = 1$. Since the 2–norm of a vector is greater than or equal to any element of that vector, no element of X_1 or X_2 can be greater than 1. Therefore, in forming $X_{new} = X_1 + (\alpha_1 + \jmath\alpha_2)X_2$, α_1 and/or α_2 values much greater than 1 will essentially eliminate any contribution of X_1 so X_{new} quickly reaches its asymptotic value of X_2. In the rare case that such a worst case condition occurs, values of α_1 or α_2 of 20 or less should be more than adequate to produce the optimum U_{new}. In practice, large values of α_1 and/or α_2 can be eliminated by merely shifting the complex multiplier from X_2 to X_1 so that

$$X_{new} = (\alpha_1 + \jmath\alpha_2)X_1 + X_2$$

For $\hat{X}_{new} \approx X_2$, the condition of $\alpha_1 = \alpha_2 = 0$ then prevents any contribution from X_1.

8.6 Principal Direction Alignment

The ability to compute candidates for μ using only $2(q-1)$ variables provides an extremely practical means for dealing with cusping singular values. After determining $\sup_{\alpha_i} \rho(M_a U_{new})$ the result may compared to the singular value upper bound and if the conservatism gap does not exceed desired tolerances no further optimization is required. The conservative upper bound is simply chosen as an estimate for μ and the frequency sweep over the desired range continues. In the event that gap between the upper and lower bounds exceeds the termination requirements, the supremizing matrix $\hat{U}_{new}$ offers an improved initial condition for a search over the full set $(n-1)$ of

optimization variables. Unfortunately, the computational aspects of this additional search are potentially overwhelming: especially as the system size increases. Rather than continuing the search for μ using eigenvalue calculations, it will be shown that a simple transformation of U_{new} allows the optimization to proceed using singular value decompositions. Because the singular values are the square roots of the eigenvalues of positive definite, symmetric matrices, ($A^H A$ or AA^H), the singular value decomposition offers significant improvements in computing efficiency over general eigenvalue decompositions. For example, using the MATLAB^{TM} "eig" and "svd" functions on 10 random, complex 9×9 matrices, the average flop counts are as follows:

$$\text{eig} \rightarrow 53{,}606 \text{ flops}$$

$$\text{svd} \rightarrow 38{,}448 \text{ flops.}$$

As the system size increases, this difference becomes even more significant indicating the obvious computational advantage of the singular value decomposition versus the eigenvalue decomposition. It must also be noted that the flop counts above include calculation of the singular vectors while the eigenvectors were not similarly determined.

The MPDA property of Chapter 4 shows that the unique stationary point associated with $\inf_D \overline{\sigma}(DM_aD^{-1})$ corresponds to a stationary point of $\sup_{U_d} \rho(M_aU_d)$. When $\overline{\sigma}$ repeats there may be no stationary point of $\overline{\sigma}$ so the direct relationship between $\overline{\sigma}$ and μ is in general lost (the "kiss" condition is an exception). However, using the principal direction alignment (PDA) property developed by Daniel, Kouvaritakis and Latchman , a similar relationship is shown to exist between $\sigma_i(DM_aD^{-1})$ and

$\rho(M_a U_d)$ where the σ_i represents some singular value other than the maximum [31]. Development of the PDA property starts with the following lemma describing the weil known eigenvalue shift property.

Lemma 8.2: For any matrix $A \in \mathcal{C}^{n\times n}$ and scalar r

$$\lambda_i(A + rI) = \lambda_i(A) + r$$

Proof: Denote the eigenvalues of matrix A as λ and of matrix $(A + rI)$ as $\hat{\lambda}$. The respective eigenvalues are simply the solutions to

$$det\{\lambda I - A\} = 0$$

and

$$det\{\hat{\lambda} I - A - rI\} = 0.$$

Now define $(\hat{\lambda} - r) = \tilde{\lambda}$ and rewrite the second equation as

$$det\{\tilde{\lambda} I - A\} = 0.$$

Since the eigenvalues of a matrix are the roots of its characteristic equation they must be unique; therefore $\tilde{\lambda} = \lambda$. Substituting back for $\hat{\lambda}$ gives $\hat{\lambda} = \lambda + r$ and thus completes the proof. ■

While the eigenvalues of a matrix obey the shift property, the singular values do not. As in Section 8.3, this provides a means of breaking up the repeated singular values by introducing a sufficiently large r. The spectral radius/singular value inequality then becomes

$$\rho(M_a U_d) + r = \rho(M_a U_d + rI) \leq \inf_D \overline{\sigma}(D M_a U_d D^{-1} + rI). \qquad (8.12)$$

Equation 8.12 requires that $\lambda_1(M_aU_d)$ corresponding to $\rho(M_aU_d)$ be real. Such a requirement is not a limitation, however, since U_d may be rotated by a scalar multiplier $e^{j\phi}$ without affecting the magnitude of $\rho(M_aU_d)$. Rewriting Equation 8.12 into a structured singular value relationship results in

$$\mu(M_a) = \sup_{U_d} \rho(M_aU_d) = \sup_{U_d} \rho(M_aU_d + rI) - r = \sup_{U_d} \inf_{D} \overline{\sigma}(DM_aU_dD^{-1} + rI) - r. \tag{8.13}$$

Unfortunately, the convexity properties that made the optimization $\inf_D \overline{\sigma}(DM_aD^{-1})$ so attractive originally are now lost because the right-hand side of Equation 8.13 is no longer invariant to U_d. Still, the addition of scalar r offers an alternative means of establishing MPDA for cusping systems. The following theorem is originally due to Latchman [27].

Theorem 8.2: Let $\overline{\sigma}(DM_aU_dD^{-1} + rI)$ have stationary points at $\hat{U}_d$ and $\hat{D}$ respectively. If r is large enough to ensure that $\overline{\sigma}(\hat{D}M_a\hat{U}_d\hat{D}^{-1} + rI)$ is simple, then MPDA is established for the matrix $(\hat{D}M_a\hat{U}_d\hat{D}^{-1} + rI)J$ where J is a matrix containing only ± 1 along the diagonal and zeros elsewhere.

Proof: The stationary points of $\sup_{U_d} \inf_D \overline{\sigma}(DM_aU_dD^{-1} + rI)$ occur when

$$\frac{\partial}{\partial d_i}[\overline{\sigma}(DM_aU_dD^{-1} + rI)] = 0 \tag{8.14}$$

and

$$\frac{\partial}{\partial \theta_i}[\overline{\sigma}(DM_aU_dD^{-1} + rI)] = 0. \tag{8.15}$$

The stationary points of Equation 8.14 have already been shown to occur upon establishment of MPDA for the matrix $(DM_aU_dD^{-1} + rI)$. This condition applies only

when

$$|X_{i1}| = |Y_{i1}| \tag{8.16}$$

for all i. Differentiation of Equation 8.15 follows in a manner similar to that of Section 5.5 so that the stationary points are characterized by

$$\frac{jr}{2}[Y_1^H E_{ii} X_1 - X_1^H E_{ii} Y_1] = 0$$

or

$$Im(X_1^H E_{ii} Y_1) = 0 \tag{8.17}$$

for all i. The requirement of Equation 8.17 may be stated in terms of the arguments of the individual vector elements such that

$$\arg(X_{i1}) = \arg(Y_{i1}) + m_i\pi \tag{8.18}$$

for all i with integer m_i. Applying Equations 8.16 and 8.18 requires that

$$X_1 = JY_1$$

and completes the proof. ■

Progressing from this point to MPDA therefore requires that $J = I$ for all cases in which both Equations 8.14 and 8.15 are satisfied. The following theorem, originally due to Latchman with simplifications by Young [27, 46] shows that J must equal I.

Theorem 8.3: Let $\overline{\sigma}(DM_aU_dD^{-1} + rI)$ have stationary points at $\hat{U}_d$ and $\hat{D}$ respectively. If r is large enough to ensure that $\overline{\sigma}(\hat{D}M_a\hat{U}_d\hat{D}^{-1} + rI)$ is simple, then MPDA is established for the matrix $(\hat{D}M_a\hat{U}_d\hat{D}^{-1} + rI)$.

Proof: As mentioned above, this proof reduces to showing that $J = I$ at all stationary points of Equation 8.13. Matrix $(DM_aU_dD^{-1} + rI)J$ has been shown to satisfy the MPDA requirement. Therefore

$$\rho(\hat{D}M_aU_d\hat{D}^{-1} + rI) \leq \overline{\sigma}([\hat{D}M_aU_d\hat{D}^{-1} + rI]J).$$

Multiplying by X_1^H and X_1 gives

$$X_1^H(\hat{D}M_aU_d\hat{D}^{-1} + rI)X_1 \leq X_1^H(\hat{D}M_aU_d\hat{D}^{-1} + rI)JX_1$$

which may be simplified as

$$rX_1^H(I - J)X_1 \leq X_1^H\hat{D}M_a\hat{D}^{-1}\hat{U}_d(J - I)X_1. \tag{8.19}$$

Clearly as r increases, the left-hand side of Equation 8.19 increases while the right-hand side remains unchanged. The inequality would therefore be contradicted unless $J = I$ which completes the proof. ■

Theorem 8.2 reveals that for r large enough to separate $\overline{\sigma}$ and σ_2

$$\sup_{U_d} \rho(M_aU_d) = \sup_{U_d} \inf_{D} \overline{\sigma}(DM_aU_dD^{-1} + rI) - r.$$

More importantly, the following lemma serves to relate stationary points of $\sup_{U_d} \inf_D \overline{\sigma}(DM_aU_dD^{-1} + rI)$ to those of $\sigma_i(DM_aU_dD^{-1})$ where the $\sigma_{i's}$ represent singular values with magnitudes less than $\overline{\sigma}$. This new formulation is invariant to U_d and no longer requires the shift by r.

Lemma 8.3: For $\hat{D}$ and $\hat{U}_d$ defined as optimizing solutions to Equation 8.13, denote X_1 as the left major singular vector of $(\hat{D}M_a\hat{U}_d\hat{D}^{-1}+rI)$. Then X_1 and its corresponding conjugate transpose, X_1^H, are also right and left eigenvectors of $(\hat{D}M_a\hat{U}_d\hat{D}^{-1})$. In

addition, matrix $(\hat{D}M_a\hat{U}_d\hat{D}^{-1})$ has some right singular vector which is aligned with its corresponding left singular vector. This alignment between nonmajor singular vectors will be referred to as Principal Direction Alignment [31].

Proof: From Theorem 8.2 it is known that $(\hat{D}M_a\hat{U}_d\hat{D}^{-1} + rI)$ has MPDA and therefore X_1 and X_1^H form a left and right eigenvector pair for $(\hat{D}M_a\hat{U}_d\hat{D}^{-1} + rI)$. From the eigenvalue shift property of Lemma 8.2, X_1 and X_1^H must form a left and right eigenvector pair for $(\hat{D}M_a\hat{U}_d\hat{D}^{-1})$ as well. This allows the eigenvalue decomposition of $(\hat{D}M_a\hat{U}_d\hat{D}^{-1})$ to be written as

$$\hat{D}M_a\hat{U}_d\hat{D}^{-1} = \left[\begin{array}{c|c} X_1 & W \end{array}\right] \left[\begin{array}{c|c} \lambda_1 & 0 \\ \hline 0 & \Lambda \end{array}\right] \left[\begin{array}{c} X_1^H \\ \hline V^T \end{array}\right]$$

where Λ, W and V^T include all the eigenvalues, and right and left eigenvectors except λ_1, X_1 and X_1^H respectively. Inspection of the hermetian products

$$(\hat{D}M_a\hat{U}_d\hat{D}^{-1})^H(\hat{D}M_a\hat{U}_d\hat{D}^{-1})$$

and

$$(\hat{D}M_a\hat{U}_d\hat{D}^{-1})(\hat{D}M_a\hat{U}_d\hat{D}^{-1})^H$$

as in Equation 8.4, reveals that X_1 is both a right and corresponding left singular vector associated with some $\sigma_i(\hat{D}M_a\hat{U}_d\hat{D}^{-1})$. Therefore, $(\hat{D}M_a\hat{U}_d\hat{D}^{-1})$ has PDA concluding the proof. ■

Combining the results from Theorem 8.2 and Lemma 8.3 leads to the following theorem concerning the direct relationship between $\sup_{U_d}\rho(M_aU_d)$ and $\sigma_i(\hat{D}M_a\hat{U}_d\hat{D}^{-1})$.

Theorem 8.4: For any matrix M_a, there always exists an optimal $\hat{U}_d$ and $\hat{D}$ such that $(\hat{D}M_a\hat{U}_d\hat{D}^{-1})$ achieves PDA and the corresponding singular value is a stationary point with respect to D. Also, under these circumstances the following condition arises:

$$\sup_{U_d} \rho(M_a U_d) = \max_i \left\{ \sigma_i(DM_aD^{-1}) \mid \frac{\partial}{\partial d_j}[\sigma_i(DM_aU_dD^{-1})] = 0 \right\} \tag{8.20}$$

for $i, j = 1, \ldots, n$.

Proof: Theorem 8.2 guarantees the existence of an optimizing $\hat{U}_d$ and $\hat{D}$ such that $(\hat{D}M_a\hat{U}_d\hat{D}^{-1})$ satisfies MPDA. The shift property may then be employed giving

$$\rho(M_a\hat{U}_d) = \overline{\sigma}(\hat{D}M_a\hat{U}_d\hat{D}^{-1} + rI) - r = \sigma_i(\hat{D}M_a\hat{D}^{-1}).$$

Theorem 8.3 shows that for optimal $\hat{D}$ and $\hat{U}_d$ matrix $(\hat{D}M_a\hat{U}_d\hat{D}^{-1})$ will have PDA. For the case of $i = 1$, $\overline{\sigma}$ does not cusp and the MPDA property is achieved. When $i > 1$, achieving PDA for $(\hat{D}M_a\hat{U}_d\hat{D}^{-1})$ requires one of the singular vector pairs to be aligned indicating that $\sigma_i(\hat{D}M_a\hat{D}^{-1})$ is a stationary point with respect to D. Selecting the largest such stationary point completes the proof. ■

Example 8.3: This example illustrates these results using the system of Table 8.3. The supremizing $\hat{U}_d$ is the same as that presented in Equation 8.11. Infimizing the right-hand side of Equation 8.20 over D with $r \geq 1.87$ gives

$$\hat{D} = diag\,[1,\ 0.749,\ 1.106,\ 1.225,\ 1.218]$$

and

$$\mu(M_a) = \rho(M_a\hat{U}_d) = \overline{\sigma}(\hat{D}M_a\hat{U}_d\hat{D}^{-1} + rI) - r = \sigma_2(\hat{D}M_a\hat{D}^{-1}) = 12.81.$$

The relationship of Equation 8.20 allows the singular value decomposition to be retained in the computation of μ even for repeated $\overline{\sigma}$. Unfortunately the stationarity condition on σ_i no longer restricts the search for μ to minima or maxima but to inflection points as well. This introduces additional complexity to an already nonconvex function as the search criteria now includes sorting through all stationary points. If, however, the search over the α_i produces a value of $\rho(M_a U_{new})$ close to $\mu(M_a)$, then an optimization based on the right-hand side of Equation 8.20 may offer tremendous computational advantages over the left-hand side of the equation.

8.7 Direct Calculation of $\hat{D}$ from $\hat{U}_d$

Section 8.6 introduced a method for directly calculating $\hat{U}_d$ given $\hat{D}$ such that $\overline{\sigma}(\hat{D} M_a \hat{D}^{-1}) = \rho(M_a \hat{U}_d) = \mu(M_a)$. Using the $2(q-1)$-dimensional search over α_i, a candidate for $\hat{U}_d$ denoted U_{new} may be calculated. In order for this estimate to serve as a useful initial guess for the right-hand side of Equation 8.20, some means of transforming U_{new} into a corresponding D_{new} must be found. Such a transformation was originally presented by Fan and Tits as a means of verifying that the solution to their nonconvex vector optimization method equals μ [38]. By modifying their original theorem and proof pertaining to this transformation it is possible to directly determine $\hat{D}$ corresponding to $\hat{U}_d$ such that $\sigma_i(\hat{D} M_a \hat{D}^{-1}) = \rho(M_a \hat{U}_d)$ even when $\overline{\sigma}$ repeats. Equation 7.2 is repeated here for convenience.

$$\mu(M_a) = \max_{x \in C^n} \{ \|M_a x\|_2 \mid \|\mathcal{P}_i x\|_2 \|M_a x\|_2 = \|\mathcal{P}_i M_a x\|_2,\ i = 1, \ldots, m \} \qquad (8.21)$$

The following proposition and subsequent theorem are originally due to Fan and Tits with modifications to account for systems with repeated singular values [38].

Proposition 8.2: Suppose $x \in C^n$ is a solution to Equation 8.21. Then there exists a vector $\hat{\lambda}_i$, $i = 1, \ldots, m$ such that the following equation holds:

$$M_a^H M_a x + \sum_{i=1}^{m} \hat{\lambda}_i \left[M_a^H \mathcal{P}_i M_a x - \|\mathcal{P}_i x\|_2^2 M_a^H M_a x - \|M_a x\|_2^2 \mathcal{P}_i x \right] = 0 \qquad (8.22)$$

where the $\mathcal{P}_i$ are projection matrices as in Chapter 7 and the $\hat{\lambda}_{i's}$ are multipliers with $\hat{\lambda}_i \geq 0$, $i = 1, \ldots, n$.

Theorem 8.5: Let $\hat{\lambda} = \left[\hat{\lambda_1}, \ldots, \hat{\lambda_n}\right]$ be the unique multiplier vector corresponding to the optimizing $\hat{x}$. Then $\hat{D} \triangleq diag\left[\hat{\lambda_1}^{\frac{1}{2}}, \ldots, \hat{\lambda_n}^{\frac{1}{2}}\right]$ is the solution to Equation 8.20.

Proof: From Chapter 7, some Y_i must exist such that

$$\hat{D}^{-1} M_a^H \hat{D}^2 M_a \hat{D}^{-1} Y_i = \sigma_i(\hat{D} M_a \hat{D}^{-1}) Y_i = \mu^2(M_a) Y_i = \|M_a x\|_2^2 Y_i.$$

Multiplying on the left by $\hat{D}$ gives the expression

$$M_a^H \hat{D}^2 M_a x = \|M_a x\|_2^2 \hat{D}^2 x. \qquad (8.23)$$

Now define some β such that

$$\beta_i = \alpha \hat{d_i}^2 \quad i = 1, \ldots, m \qquad (8.24)$$

with α selected such that $\sum_{i=1}^{m} \beta_i \|\mathcal{P}_i x\|_2^2 = 1$. Then $\alpha \hat{D}^2 = \sum_{i=1}^{m} \beta_i \mathcal{P}_i$ and Equation 8.23 becomes

$$\sum_{i=1}^{m} \beta_i \left[M_a^H \mathcal{P}_i M_a x - \|M_a x\|_2^2 \mathcal{P}_i x \right] = 0$$

giving

$$M_a^H M_a x + \sum_{i=1}^{n} \beta_i \left[M_a^H \mathcal{P}_i M_a x - \|\mathcal{P}_i x\|_2^2 M_a^H M_a x - \|M_a x\|_2^2 \mathcal{P}_i x \right] = 0.$$

Equating $\beta_i = \hat{\lambda}_i$, this equation becomes identical to Equation 8.22 and through the relationship of Equation 8.24 the proof is complete. ■

It is now possible to directly determine a $\hat{D}$ given $\hat{U}_d$ such that $\rho(M_a\hat{U}_d) = \sigma_i(\hat{D}M_a\hat{D}^{-1}) = \mu(M_a)$. Again using $\hat{U}_d$ from Equation 8.11, $\hat{D}$ is found to be

$$\hat{D} = diag\,[1,\ 0.749,\ 1.106,\ 1.225,\ 1.218]$$

with

$$\mu(M_a) = \rho(M_a\hat{U}_d) = \sigma_2(\hat{D}M_a\hat{D}^{-1}) = 12.81$$

providing the same results as before. However, no additional optimization is required for this method as opposed to Example 8.3 which requires an infimization over D of $\overline{\sigma}(DM_aU_dD^{-1} + rI)$. Provided that the $2(q-1)$-dimensional optimization generates a U_{new} close to $\hat{U}_d$, the transformation from U_{new} to D_{new} may be invoked to solve for μ using the method of Equation 8.20.

8.8 Comparison of Optimization Methods

The results of this chapter are summarized in Tables 8.6 and 8.7. Table 8.6 lists system size, the results of $\inf_D \overline{\sigma}(DM_aD^{-1})$, the value $\mu(M_a)$, the results of $\sup_{\alpha_i} \rho(M_aU_{new})$, and the per cent difference between $\mu(M_a)$ and $\sup_{\alpha_i} \rho(M_aU_{new})$. System 6 corresponds to that of Figure 8.3 and, as mentioned earlier, it shows the largest deviation between $\mu(M_a)$ and $\sup_{\alpha_i} \rho(M_aU_{new})$ of all cusping systems studied to date. System 5 shows the largest deviation between $\inf_D \overline{\sigma}(DM_aD^{-1})$ and $\mu(M_a)$ while the corresponding value of $\sup_{\alpha_i} \rho(M_aU_{new})$ equals $\mu(M_a)$ within four significant

figures. Clearly the result of the $2(q-1)$-dimensional optimization $\sup_{\alpha_i} \rho(M_a U_{new})$ provides an excellent estimate of μ.

Depending on the particular control system under development, it is likely that these estimates of μ will be satisfactory without any additional optimization. In the event that μ must be determined exactly, the optimum $\hat{U}_{new}$ from $\sup_{\alpha_i} \rho(M_a U_{new})$ may be used as an starting point for further searches. Using the same systems as Table 8.6, Table 8.7 summarizes the flop count requirements of the various methods. Each of the optimization methods call a series of MATLAB^{TM} m-files developed by Grace [33]. While m-files generally execute more slowly than compiled code, the ability to track flop counts provides a means of comparing the efficiency of algorithms regardless of the host machine.

In order to provide as equitable a comparison between methods as possible each cusping system was required to satisfy three criteria. Of all cusping systems tested, only those meeting these criteria are included in Tables 8.6 and 8.7. (Note: It should be mentioned that failure to satisfy these criteria does not mean that the optimization methods could not find μ for a particular system, it simply means that any flop count comparison would be erroneous.)

The first selection criterion is that optimization methods A and B ($\sup_{U_d} \rho(M_a U_d)$ and $\sup_{\alpha_i} \rho(M_a U_{new})$ respectively), must begin at the same starting point. This point is a value of U_d generated by the method of Theorem 8.1. For $q = 2$, two values of U_d are generated with U_1 corresponding to X_1 and Y_1 and U_2 corresponding to X_2 and Y_2. Since these singular vector pairs are effectively interchangeable at the cusp

Table 8.6: Cusping Systems

System	Size	$\inf_{\mathcal{D}} \overline{\sigma}(DM_aD^{-1})$	μ	$\sup_{\alpha_i} \rho(M_aU_{new})$	% Difference μ vs $\rho(M_aU_{new})$
1	4	6.8125	6.7803	6.7803	-0.000
2	4	5.9398	5.9015	5.9015	-0.000
3	4	7.1238	6.9697	6.9488	-0.013
4	4	6.5750	6.3673	6.3671	-0.003
5	4	1.0322	0.9470	0.9470	-0.000
6	5	13.114	12.810	12.761	-0.326
7	5	31.710	31.443	31.440	-0.009
8	5	35.838	35.615	35.604	-0.031
9	5	43.913	43.809	43.807	-0.007
10	5	35.126	34.930	34.922	-0.022
11	9	7.7400	7.7394	7.7394	-0.000
12	9	7.8366	7.8346	7.8346	-0.000
13	9	9.7072	9.6193	9.6185	-0.008
14	9	29.473	29.254	29.253	-0.003
15	9	128.02	127.07	127.07	-0.002
16	9	74.749	74.502	74.501	-0.001
17	9	151.95	150.29	150.28	-0.002
18	9	209.64	209.21	209.20	-0.001
19	9	35.238	34.610	34.610	-0.001
20	9	35.918	35.339	35.336	-0.008

Table 8.7: Floating Point Operations Comparison

	Optimization Method			
	A	B	C	D
System	$\sup_{U_d} \rho(M_a U_d)$	$\sup_{\alpha_i} \rho(M_a U_{new})$	$B \downarrow$ $\sup_{U_d} \rho(M_a U_d)$	$B \downarrow$ $\max_i \sigma_i(DM_a D^{-1})$
1	234,427	128,578	149,067	149,899
2	138,177	113,174	144,883	150,354
3	128,398	136,015	191,239	157,372
4	188,109	193,382	212,596	214,373
5	171,721	63,162	74,317	79,787
Average:	172,166	126,862	154,420	150,357
6	579,929	301,294	395,592	343,173
7	382,910	299,005	550,583	343,313
8	709,431	285,912	452,255	329,634
9	839,772	553,011	722,400	595,733
10	415,088	256,610	582,077	300,868
Average:	585,426	339,166	540,581	382,544
11	677,513	226,805	408,046	452,390
12	1,037,283	384,400	563,199	611,346
13	1,992,286	816,698	1,221,836	1,039,923
14	2,909,386	1,172,394	1,850,040	1,398,542
15	2,850,485	1,078,454	2,345,768	1,307,238
16	3,203,488	984,491	1,952,090	1,211,701
17	3,327,218	1,240,607	2,200,240	1,465,977
18	2,786,795	1,330,380	2,342,328	1,556,346
19	3,349,508	946,319	1,687,675	1,113,345
20	4,492,976	1,684,968	3,521,077	1,937,664
Average:	2,662,694	986,551	1,809,229	1,109,472

(see Equations 8.4 and 8.5), both U_1 and U_2 provide candidate starting points for the subsequent optimizations. The choice between the two is made by selecting the maximizer of $\rho(M_a U_i)$ for $i = 1, \ldots, q$. The extra flops required to select the desired U_i are generally more than offset by increased convergence rates resulting from the superior starting point.

Choosing U_1 or U_2 as a starting point for $\sup_{U_d} \rho(M_a U_d)$ corresponds to choosing whether the complex multiplier $(\alpha_1 + \jmath\alpha_2)$ acts on X_1 or X_2 in Equation 8.7. Using $\alpha_1 = \alpha_2 = 0$ as an initial condition, matrix U_1 results from

$$X_{new} = X_1 + (\alpha_1 + \jmath\alpha_2)X_2$$

while U_2 results from

$$X_{new} = (\alpha_1 + \jmath\alpha_2)X_1 + X_2.$$

This procedure allows $\sup_{U_d} \rho(M_a U_d)$ and $\sup_{\alpha_i} \rho(M_a U_{new})$ to start at the same point satisfying the first selection criterion.

The second criterion requires the optimization to converge to μ from the common starting point. This greatly reduces the number of systems available for flop count comparisons because the methods are all known to be nonconvex. While changing the initial starting points would eventually lead to μ, the possibility of an equitable flop count comparison is eliminated. This local convexity criterion extends to the methods of columns C and D of Table 8.7 as well. For these methods, $\sup_{\alpha_i} \rho(M_a U_{new})$ provides an improved starting point for subsequent optimizations by $\sup_{U_d} \rho(M_a U_d)$ and $\max_i \sigma_i(DM_a D^{-1})$ respectively. As discussed in Section 8.6, the optimization based on $\sigma_i(DM_a D^{-1})$ searches for all stationary points including maxima, minima,

and inflection points. While the computational advantages of the singular value decomposition makes it quite fast, its nonconvexity generally prevents its use without the improved starting point provided by $\sup_{\alpha_i} \rho(M_a U_{new})$.

The final selection criterion involves termination conditions for the separate methods. While the end results of methods A, C, and D should be identical, differences in problem formulation prevent the use of a common termination condition. If one algorithm employs a more strict termination condition, it may unnecessarily require more flops than a second, less strict method to achieve the same results. To resolve this dilemma, parameters were adjusted to halt the optimization of $\sup_{U_d} \rho(M_a U_d)$ when $\|\frac{\partial \rho}{\partial \theta_i}\| < 0.02$. Corresponding parameters for $\max_i \sigma_i(DM_aD^{-1})$ were then set such that termination occurred when $\max_i \sigma_i(DM_aD^{-1}) = \rho(M_a\hat{U}_d)$ within four significant figures.

Termination conditions for $\sup_{\alpha_i} \rho(M_a U_{new})$ were adjusted such that the optimization successfully halts when all terms of the numerical gradient fall below 10^{-2}. Tests with varying termination conditions suggest that further reductions in the numerical gradient produce little if any improvement in $\rho(M_a U_{new})$.

Having discussed the three selection criteria, the flop count comparison of Table 8.7 may be examined. The superiority of Method B in terms of computing efficiency is immediately apparent. For the 4×4 systems, an average of 172,166 flops was required by Method A versus 126,862 flops for Method B: a savings of 26%. This savings increases to 42% for the 5×5 systems and 63% for the 9×9 systems. For

design problems requiring exact values of μ, Method D offers a savings of 58% over Method A and 39% over Method C for 9×9 systems.

The efficiency of Method D for finding the exact value of μ seems to be a direct result of the excellent starting point provided by first solving for $\sup_{\alpha_i} \rho(M_a U_{new})$. In fact, for many of the 20 systems the flop count difference between Methods B and D simply reflects the calculations involved in computing $\frac{\partial \sigma_i}{\partial d_i}$ one time. The resulting derivative was sufficiently small to meet the termination criteria without further optimization.

It is important to emphasize that the criteria chosen for the flop count comparison, especially the second criterion, actually skew the results against the new $2(q-1)$−dimensional search method. While the flop count savings are quite substantial, the real advantages occur for systems that do not directly converge to μ from the starting point. Convergence to some local maxima requires the selection of a new starting point from which the search for μ may continue. For previous methods this means a search over $n-1$ dimensions so the search area grows exponentially as system size increases. The new method, however, requires a search over only a $2(q-1)$ dimensional space greatly increasing the likelihood of quickly finding μ.

8.9 Algorithm for Cusping Systems

Certainly the numerical efficiency of $\sup_{\alpha_i} \rho(M_a U_{new})$ offers advantages even if the exact value of μ must be found. For the general cusping case of $q = 2$, the ability to generate and view 3-D mesh plots as in Figure 8.4 offers a means of partially overcoming the problem of nonconvexity. A general algorithm for computing μ to

account for cusping and noncusping systems is now presented.

1. Compute $\inf_D \overline{\sigma}(DM_aD^{-1})$. If MPDA is established then $\overline{\sigma}(\hat{D}M_a\hat{D}^{-1}) = \mu(M_a)$ and the algorithm is complete.

2. If MPDA is not established then no stationary point was found and the system is either a "kiss" or a "cusp." Determine the first q singular vectors of (DM_aD^{-1}) where q is the multiplicity of $\overline{\sigma}$ and compute $\sup_{\alpha_i} \rho(M_aU_{new})$. If MPDA is now established ("kiss") or $\rho(M_a\hat{U}_{new})$ is within some allowable distance of $\overline{\sigma}(DM_aD^{-1})$ at the cusp then the procedure is complete and upper bound $\overline{\sigma}(DM_aD^{-1})$ is used for $\mu(M_a)$.

3. If $\rho(M_a\hat{U}_{new})$ falls below the desirable distance from $\overline{\sigma}(DM_aD^{-1})$ then continue with the PDA search over $\sigma_i(DM_aD^{-1})$ (Method D of Table 8.7) to determine a true stationary point.

4. If the stationary point from Step 3 is still not acceptable then either change the initial settings of the α_i and repeat Steps 2 and 3 or generate a mesh of $\rho(M_aU_{new})$ over the α_i as discussed in Section 8.4. Then investigate all local maxima from the mesh for the global maximum of $\sup_{U_d} \rho(M_aU_d)$.

While even the mesh is not guaranteed to contain μ, the advantages offered by the optimization $\sup_{\alpha_i} \rho(M_aU_{new})$ are a tremendous improvement over previous methods of computing μ for cusping systems. It should also be mentioned that this approach for computing μ also extends to cusping block structured systems. Equation 5.10 reveals the relationships between the elements of $X_1, \ldots, X_q$ and $Y_1, \ldots, Y_q$ which

must hold even when cusping occurs. Similar relationships can be determined for any general block structured system so forming X_{new} as linear combinations of $X_1, \ldots, X_q$ proceeds exactly as before. The only difference between the element-by-element uncertainty algorithm and that for general block structured systems is in the formulation of U_{new} which must reflect the associated block configuration. Otherwise, the same reduction in optimization variables with corresponding efficiency benefits applies.

CHAPTER 9
CONCLUSION

9.1 Summary

This dissertation presents new results in the area of stability robustness analysis for multivariable systems. Of primary significance are the links developed between similarity and nonsimilarity scaling techniques for computing the structured singular value (μ). The results of Chapter 5 show that no more than $2(n-1)$ optimization variables are required to compute μ using either similarity or nonsimilarity scaling. While the nonsimilarity scaling approach effectively addresses element-by-element structured uncertainties, the ability of similarity scaling to handle general block structured uncertainties with no more than $2(n-1)$ free variables greatly simplifies the calculation of μ for this important uncertainty class.

For the general block 2×2 problem, the elimination of redundant variables leads to a guarantee that the solution of $\inf_{D_b} \overline{\sigma}(D_b M_b D_b^{-1})$ equals $\mu(M_b)$ regardless of whether or not a stationary point occurs at the "inf." This eliminates the need to employ one of the computationally intensive lower bound optimization routines for this class in the event that the maximum singular value repeats.

In addition to the reduction in optimization variables, the link between the two scaling methods is shown in Chapter 6 to include a direct relationship between scaling matrix D of similarity scaling and scaling matrices R and L of nonsimilarity scaling.

The value of this direct relationship is shown to be most evident when applied to methods that compute a lower bound for μ as described in Chapter 7. Since such methods involve nonconvex optimizations, their solutions are initially only candidates for μ. Guaranteeing that some solution actually equals μ requires showing that the candidate and the upper bound of $\overline{\sigma}(\hat{D}M_a\hat{D}^{-1})$ are equal. Computing the corresponding $\hat{D}$ for a particular μ candidate is straightforward; however, the actual singular value decomposition must then be performed on D and M_a matrices with dimensions as large as $n^2 \times n^2$. By directly transforming $\hat{D}$ to the corresponding $\hat{R}$ and $\hat{L}$, the upper bound may be determined using the $n \times n$ decomposition $\overline{\sigma}(\hat{R}^{-1}M\hat{L}^{-1})\overline{\sigma}(\hat{L}P\hat{R})$ with a substantial reduction in floating point operations for each step in the required frequency sweep.

One of the most important advantages of computing the singular value upper bound for μ is that any local minimum must be the global minimum. As long as this minimum corresponds to a stationary point, the MPDA theory guarantees that this upper bound equals μ. However, for those cases that do not achieve stationarity at the minimum ("cusps"), μ is not achieved and, in fact, the actual value of μ may differ from the upper bound by more than 15%. Because of the nonconvexity involved in lower bound calculations and the dangers of choosing an optimistic estimate for μ, the conservative upper bound is generally used in place of μ when cusping occurs. This is particularly true as system size increases because the lower bound solution domain increases directly with n, greatly impeding the search process.

By applying the MPDA principal, an estimate for μ requiring only $2(q-1)$ op-

timization variables (where q denotes the multiplicity of the cusp) is proposed in Chapter 8. The largest deviation found to date between this estimate and the actual value of μ is less than 0.34%: well within most real-world engineering tolerances. Reductions in floating point operations of more than 60% for 9×9 systems are experienced versus previous methods and for the most common cusping case of $q = 2$, 3–dimensional mesh plots may be used to easily locate candidates for μ. The μ estimate holds for cusping systems with element-by-element as well as block structured uncertainties so any system with structured uncertainties can benefit from this method. Also, the estimate can serve as an initial st rting point in the event that μ must be determined exactly. Such a procedure employing the Principal Direction Alignment theorem is shown in Table 8.6 to find the actual value of μ with only a slight increase in floating point operations over that required for the estimate.

9.2 Future Directions

9.2.1 Nonsimilarity Scaling with Block Structured Uncertainties

The results of this dissertation serve to enhance many of the important tools used to analyze the robust stability properties of multivariable control systems. In the process of completing one particular objective, other promising areas frequently appear as directions for additional research. One of the more intriguing of such areas emerging from this work is the possibility of adapting the nonsimilarity scaling technique to work with general block structured uncertainties.

For element-by-element structured uncertainties, the advantages of nonsimilarity scaling are quite clear. For example, using the similarity scaling approach, System

12 of Table 5.2 required 922,383 flops to compute μ using the reduced scaling structure. For the same system, the nonsimilarity scaling approach required only 105,061 flops: a reduction of more than 88%. Unfortunately, current nonsimilarity scaling formulations do not treat general block structured uncertainties so the tremendous advantages cannot be realized for this important uncertainty class.

While uncertainty blocks with arbitrary dimension currently prevent the use of nonsimilarity scaling methods, the direct relationship between D, R, and L of Chapter 6 may be applied to at least one specific block structure. This structure requires only that all the blocks be of a common size so that more than the standard scalar uncertainty elements are allowed. To illustrate the application of nonsimilarity scaling to this class of uncertainties, consider the system of Example 5.4 on page 61. A direct calculation of $\hat{R}$, and $\hat{L}$ from $\hat{D}$ gives

$$\hat{R} = diag[1,\ 1,\ 0.833,\ 0.833], \quad \hat{L} = diag[1,\ 1,\ 0.625,\ 0.625]$$

with $\overline{\sigma}(\hat{R}^{-1}M\hat{L}^{-1})\overline{\sigma}(\hat{L}P\hat{R}) = \overline{\sigma}(\hat{D}_b M_b \hat{D}_b^{-1}) = 16.43 = \mu(M_b)$. Since the direct transformation holds here, the problem could have been solved by employing nonsimilarity scaling from the start. The ability of nonsimilarity scaling to address even this limited class of block structured uncertainties combined with the correspondence between optimization variable requirements suggests that some general block uncertainty form of nonsimilarity scaling may eventually be possible.

9.2.2 H^∞ and μ–Synthesis Design Methods

Two recent advances in the design of multivariable control systems are the H^∞ and μ–synthesis methods. The H^∞ technique allows for the design of controllers

that possess optimal performance in terms of minimizing the H^∞ norm of some frequency dependent transfer matrix [18]. On the other hand, μ–synthesis extends the optimal performance properties of the H^∞ method to encompass robust performance characteristics as well [19]. A number of interesting design studies involving the μ–synthesis approach have appeared including a flexible antenna structure and an implementation of the space shuttle lateral axis flight control system [47, 48]. The next few paragraphs give a brief overview of these techniques along with possible applications of the results of this dissertation.

For transfer matrix G, the H^∞ norm of G may be written as

$$\|G(s)\|_\infty = \sup_s \overline{\sigma}(G(s)) \tag{9.1}$$

where s again denotes frequency dependence [49]. Figure 9.1 depicts the standard H^∞ system model with feedback controller K. Transfer matrix G can be partitioned as

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \tag{9.2}$$

giving a set of system equations of the form

$$z = G_{11}w + G_{12}u, \quad y = G_{21}w + G_{22}u.$$

Noting that $u = Ky$ allows this to be rewritten as

$$z = [G_{11} + G_{12}K(I - G_{22}K)^{-1}G_{21}]w \tag{9.3}$$

Defining $F = [G_{11} + G_{12}K(I - G_{22}K)^{-1}G_{21}]$ as the transfer matrix from external input, w, to output z, Equation 9.3 can then be written as

$$z = Fw.$$

The design objective then involves minimizing the H^∞ norm of F. Depending the on the performance objective(s) to be optimized, the choice of K that best meets the objective(s) can be chosen from the set of all stabilizing controllers. Characterization of all proper, stabilizing controllers can be obtained from the so-called "Youla parameterization" [50] allowing the minimization to be formulated as

$$\min_{\text{stable } K} \sup_{s} \overline{\sigma}(F(s)) = \min_{\text{stable } K} \|F(s)\|_\infty. \tag{9.4}$$

This problem turns out to be convex and a method to directly solve for the optimal controller has been proposed by Doyle et al [51].

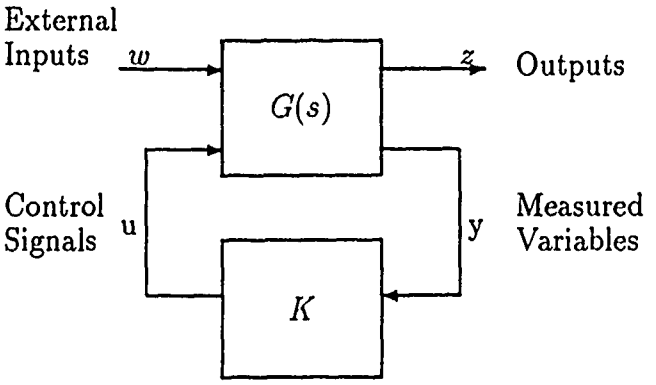

Figure 9.1: H^∞ System Representation.

The optimal performance offered by a system satisfying Equation 9.4 depends on the adequacy of the plant model. Uncertainty in this model may lead to unpredictable behavior that quickly loses optimal and even acceptable performance levels. However, by combining the H^∞ design procedure with singular value scaling techniques, robust performance and stability objectives may be simultaneously met. Such a methodology was introduced by Doyle under the name "μ–synthesis" [19]. Rather

than simply minimizing $\overline{\sigma}(F(s))$ as a function of frequency and stabilizing controller, K, $\mu-$synthesis allows for the incorporation of structured plant uncertainties with the design objective of solving

$$\min_{\text{stable } K} \inf_{D} \|D\tilde{F}D^{-1}\|_{\infty} \tag{9.5}$$

where $\tilde{F}$ represents the appropriate combination of F and Δ. For fixed D, Equation 9.5 is a standard H^{∞} problem. Similarly for fixed K the problem involves finding $\mu(D\tilde{F}D^{-1})$. Separately, each problem is convex. Unfortunately, the combined problem loses this desirable convexity property so there is no guarantee of ever achieving the global minimum in terms of both K and D.

While the $\mu-$synthesis approach provides controllers with robust performance, it requires a tremendous amount of computing power to find even a local minimum of Equation 9.5. With the loss of convexity, computational requirements increase even more. However, by applying the results of this dissertation, it may be possible to improve the efficiency of the $\mu-$synthesis method.

For example, the reduction of optimization variables for systems with both scalar and block structured uncertainties should directly apply to that step of the procedure involved in infimizing the maximum singular value over D. Also, the direct relationship between similarity scaling matrix, D, and nonsimilarity scaling matrices L and R might provide insight into a formulation of $\mu-$synthesis involving only nonsimilarity scaling with its corresponding reduced system size. Finally, the procedure outlined in Chapter 8 for estimating μ with a repeated maximum singular value should apply to the $\mu-$synthesis problem when cusping occurs so that unnecessary conservatism

can be reduced without the huge computing cost involved with previous methods of determining the lower bound for μ.

Although the actual benefits of applying the results of this dissertation to the μ–synthesis design approach are as yet unknown, any efficiency advantage over previous methods is certainly welcome. Since all but the simplest of design problems require tradeoffs between performance requirements, an iterative process is normally employed to achieve a satisfactory performance compromise. Efficiency improvements as low as ten or twenty percent could therefore translate into several hours worth of CPU savings as an acceptable performance trade-off is reached.

REFERENCES

1. L. Ljung, System Identification, Prentice-Hall, Englewood Cliffs, NJ, 1987.

2. B. Friedland, Control System Design, McGraw-Hill, New York, 1986.

3. A. Yue and I. Postlethwaite, "Improvement of Helicopter Handling Qualities Using H^{∞}−Optimisation," Proc. IEE, vol. 137, Pt. D, no. 3, May 1990, pp. 115-130.

4. F. Neslinc and S. Joshi, "Aerospace Control Systems," IEEE Control Systems Magazine, vol. 10, no. 4, June 1990, pp. 3-4.

5. M. Morari, and E. Zafiriou, Robust Process Control, Prentice Hall, Englewood Cliffs, NJ, 1989.

6. H. S. Black, "Stabilized Feed-Back Amplifiers," Electrical Engineering, vol. 53, Jan. 1934, pp. 114-120.

7. H. S. Black, "Inventing the Negative Feedback Amplifier," IEEE Spectrum, Dec. 1977, pp. 55-60.

8. H. Nyquist, "Regeneration Theory," Bell Systems Tech. J., vol. 11, 1932, pp. 126-147.

9. H. Bode, Network Analysis and Feedback Amplifier Design, Van Nostrand, New York, 1945.

10. P. Dorato, "Robust Control: A Historical Review," Proc. IEEE 25^{th} CDC., 1986, pp. 346-349.

11. A. MacFarlane, and I. Postlethwaite, "The Generalised Nyquist Stability Criterion and Multivariable Root-Loci", Int. J. Control, vol. 25, 1977, pp. 81-127.

12. J. Doyle and G. Stein, "Multivariable Feedback Design: Concept for a Classical/Modern Synthesis," IEEE Trans AC, vol. AC-26, 1981, pp. 4-16.

13. "Challenges to Control: A Collective View," IEEE Trans AC, vol. AC-32, 1987, pp. 275-285.

14. M. Safonov, "Stability Margins of Diagonally Perturbed Multivariable Feedback Systems," IEE Proc., vol. 129, Pt. D, no. 6, Nov. 1982, pp. 251- 256.

15. J. Doyle, "Analysis of Feedback Systems with Structured Uncertainties," Proc. IEE, vol. 129, Pt. D, no. 6, Nov. 1982, pp. 242-250.

16. B. Kouvaritakis and H. Latchman, "Singular Value and Eigenvalue Techniques in the Analysis of Systems with Structured Perturbations," Int. J. Control, vol. 41, no. 6, 1985, pp. 1381-1412.

17. B. Kouvaritakis and H. Latchman, "Necessary and Sufficient Stability Criterion for Systems with Structured Uncertainties: The Major Principal Direction Alignment Principle," Int. J. Control, vol. 42, no. 3, 1985, pp. 575-598.

18. G. Zames, "Feedback and Optimal Sensitivity: Model Reference Transformations, Multiplicative Seminorms, and Approximate Inverses," IEEE Trans AC, vol. AC-26, 1981, pp. 301-320.

19. J. Doyle, "Structured Uncertainty in Control System Design," Proc. 24^{th} CDC., Ft Lauderdale, FL, Dec. 1985, pp. 260-265.

20. Y. Foo, "Robustness of Multivariable Feedback Systems: Analysis and Optimal Design," D. Phil. Thesis, Oxford University, Oxford, England, 1985.

21. J. M. Maciejowski, Multivariable Feedback Design, Addison-Wesley Publishing Company, Wokingham England, 1989.

22. R. Smith and J. Doyle, "Model Invalidation: A Connection between Robust Control and Identification," Proc. ACC, 1989, pp. 1435-1440.

23. M. Fan, A. Tits and J. Doyle, "Robustness in the Presence of Joint Parametric Uncertainty and Unmodeled Dynamics," Proc. ACC, 1988, pp. 1195-1200.

24. R. De Gaston and M. Safonov, "Exact Calculation of the Multiloop Stability Margin," IEEE Trans AC, vol. AC-33, no. 2, Feb. 1988, pp. 156-171.

25. R. Peña and A. Sideris, "Robustness with Real Parametric and Structured Complex Uncertainty," Proc. IEEE 27^{th} CDC., 1988, pp. 532-537.

26. P. Young and J. Doyle, "Computation of the Structured Singular Value with Real and Complex Uncertainties," Proc. IEEE 29^{th} CDC., 1990.

27. H. Latchman, "Frequency Response Methods for Uncertain Multivariable Systems", D. Phil. Thesis, Oxford University, 1986.

28. M Safonov and J Doyle, "Optimal Scaling for Multivariable Stability Margin Singular Value Computation," MECO/EES '83 Symposium, Athens, Greece, Aug. 29-Sep. 2, 1983, pp. 466-469.

29. R. Sezginer and M. Overton, "The Largest Singular Value of $e^X A_o e^{-X}$ is Convex on Convex Sets of Commuting Matrices," IEEE Trans AC, vol. AC-vol. 35, no. 2, Feb. 1990, pp. 229-230.

30. E. Osborne, "On Pre-Conditioning of Matrices," J. Assoc. Comput. Mach., 1960, vol. 7, pp. 338-345.

31. R. Daniel, B. Kouvaritakis, and H. Latchman, "Principal Direction Alignment: A Geometric Framework for the Complete Solution to the μ-Problem," IEE Proc., vol. 133, Pt. D, no. 2, Mar. 1986, pp. 45-56.

32. PC-MATLAB User's Guide, The MathWorks Inc., South Natick, MA, 1989.

33. A. Grace, "Computer-Aided Control System Design Using Optimization Techniques," Ph.D. Thesis, University of Wales, Bangor, England, 1989.

34. W. Press, Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, New York, NY, 1986.

35. J. Doyle and C. Chu, "Robustness Control of Multivariable and Large Scale Systems," Final Technical Report for AFOSR, Contract F49620-84-C-0088, Mar. 1986.

36. N. Tsing, "Convexity of the Largest Singular Value of $e^D M e^{-D}$: A Convexity Lemma," IEEE Trans AC, vol. AC-35, no. 6, June 1990, pp. 748-749.

37. M. K. H. Fan and A. L. Tits, "A New Formula for the Structured Singular Value," Proc. IEEE 24^{th} CDC., 1985, pp. 595-596.

38. M. K. H. Fan and A. L. Tits, "Characterization and Efficient Computation of the Structured Singular Value," IEEE Trans AC, vol. AC- 31, no. 8, pp. 734-743, Aug. 1986.

39. "Harwell Subroutine Library," Library Reference Manual, Harwell, England, 1982.

40. The μ-Analysis and Synthesis Toolbox, The MathWorks Inc., South Natick, MA, 1990.

41. A. Packard, M. Fan, and J. Doyle, "A Power Method for the Structured Singular Value," Proc. IEEE 27^{th} CDC, 1988, pp. 2132-2137.

42. G. Golub and C. Van Loan, Matrix Computations, The John Hopkins University Press, Baltimore, Maryland, 1983.

43. A. Packard and J. Doyle, "Structured Singular Value with Repeated Scalar Blocks," Proc. ACC, 1988, pp. 1213-1218.

44. Y. Hung, A. MacFarlane, Multivariable Feedback: A Quasi-Classical Approach, Springer-Verlag, Berlin, 1982.

45. G. Stewart, "Error and Perturbation Bounds for Subspaces Associated with Certain Eigenvalue Problems," SIAM Review, vol. 15, Oct., 1973, pp. 727-764.

46. P. Young, "Frequency Domain Analysis of Multivariable Systems with Structured Uncertainties," Master's Thesis, University of Florida, 1988.

47. B. Balas, C. Chu, J. Doyle, "Vibration Damping and Robust Control of the JPL/AFAL Experiment Using μ–Synthesis," Proc. IEEE 28^{th} CDC., 1989, pp. 2689-2694.

48. J. Doyle, K. Lenz, and A. Packard "Design Examples Using μ-Synthesis: Space Shuttle Lateral Axis FCS During Reentry," Modelling, Robustness and Sensitivity Reduction in Control Systems, Springer-Verlag, Berlin, 1987.

49. B. Francis, A Course in H_∞ Control Theory, Springer-Verlag, Berlin, 1987.

50. D. Youla, H. Jabr, and J. Bongiorno, "Modern Weiner-Hopf Design of Optimal Controllers, Part II: The Multivariable Case," IEEE Trans AC, vol. AC-21, pp. 319-338, 1976.

51. J. Doyle, K. Glover, P. Khargonekar, and B. Francis, "State-Space Solutions to Standard H_2 and H-Infinity Control Problems," Proc. ACC, 1988, pp. 1691-1696.

BIOGRAPHICAL SKETCH

Robert Jeffrey Norris was born on October 13, 1959, in Morehead City, North Carolina. He graduated summa cum laude from North Carolina State University in 1981 with a degree in electrical engineering. He then entered graduate school at North Carolina State University and received a master's degree in electrical engineering in 1983. After completing his master's degree he received a commission in the United States Air Force where his work involved the guidance and control of remotely piloted vehicles. In 1987 he was selected for an Air Force fellowship to study electrical engineering at the University of Florida.